TI 2018-010/I
Tinbergen Institute Discussion Paper

The Dark Side of Personality: Anti-Sociality Increases Strategic Game Play

Jan (J.B.) Engelmann[1]
Basil Schmid[2]
Justin Chumbley[3]
Ernst Fehr[4]

1: University of Amsterdam; Tinbergen Institute, The Netherlands
2: ETH Zurich
3: ETH Zurich
4: University of Zurich

The Dark Side of Personality: Anti-Sociality Increases Strategic Game Play

Engelmann, J. B.[1,2*], Schmid, B.[3], Chumbley, J.[4] & Fehr, E.[5]

1) Center for Research in Experimental Economics and Political Decision Making (CREED), University of Amsterdam and the Tinbergen Institute, Amsterdam, The Netherlands
2) Amsterdam Brain and Cognition (ABC), University of Amsterdam, Amsterdam, The Netherlands
3) Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich, Switzerland
4) Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland
5) Department of Economics, University of Zurich, Zurich, Switzerland

*Please address correspondence to:

j.b.engelmann@uva.nl

Abstract

We assess the role of anti-social personality traits in explaining heterogeneity in commonly observed social preferences. We identified a personality profile that clearly reflects anti-social personality characteristics, with high positive loadings on Machiavellianism and high negative loadings on empathy, trustworthiness and agreeableness. Anti-sociality predicts decision strategies in a manner that is consistent with its name: significantly lower levels of trust and decreased trustworthiness. To identify the strategic nature of anti-social behavior in changing environments, we assessed the moderating role of personality on investor trust and trustee reciprocity in the presence relative to the absence of the investor's option to punish. Our results show that only the anti-social personality profile is associated with specific payoff maximizing strategy shifts induced by these environmental changes: when punishment was not available to investors, we observe significantly lower levels of investor trust and trustee reciprocity, while there is a significant increase in both behaviors when punishment was available. These effects were specific for anti-sociality, as no other personality factor was associated with such a strong adjustment of decision strategies in the presence of punishment. These results demonstrate that anti-social personality characteristics are associated with strategic behavioral shifts aimed at maximizing the extraction of resources from their counterparts. The reliability of the strategic effects of anti-social personality during trust, reciprocity and punishment strongly supports the notion that self-projection underlies anti-social decision-making.

JEL: C71, D87, D91

Keywords: trust, reciprocity, punishment, anti-social, personality, individual differences

/body

Introduction

Previous research has uncovered the presence of substantial heterogeneity in social preferences that include for instance reciprocal fairness (Falk and Fischbacher, 1999), inequity aversion (Fehr and Schmidt, 1999), and altruism (Fehr and Fischbacher, 2002), in addition to purely selfish and even spiteful behavioral tendencies (Falk et al., 2000). A key question that remains unanswered to date concerns the sources that generate such heterogeneity in social preferences. Understanding the origins of such heterogeneity can provide important insights into the mechanisms that drive choices, assist in developing better models for predicting behavior within specific environments and has important implications for policy analysis (e.g., Heckman, 2001; Dohmen et al., 2008). Here, we integrate approaches from two disciplines to investigate the sources of individual differences in other-regarding preferences: Behavioral Economics and Personality Psychology. The intersection between Personality Psychology and Behavioral Economics has experienced emerging interest in the recent past (Capra et al., 2013; Becker et al., 2012; Borghans et al., 2008, Almlund et al., 2011, Ferguson et al., 2011, Zhao and Smilie, 2015). Personality traits measure the relatively enduring patterns of thoughts, emotions and behavior that reflect an individual's propensity to respond in certain ways under specific circumstances (Roberts, 2009), thereby providing a set of coherent constructs that are complementary to economic preference measures (Becker et al., 2012). Personality traits could therefore help explain the complexities in manifestations of economic and social behavior (Ferguson et al., 2011).

Results from several investigations that have employed this approach underline the notion that personality is an important factor that mediates the interplay between cognition and emotion in the production of behavior. Specifically, conscientiousness and emotional stability predict performance on tests commonly regarded as measuring purely cognitive abilities, such as intelligence (Borghans et al., 2008) and academic achievement (Cunha and Heckman, 2008; Cunha et al., 2010). Further evidence suggests that personality has a specific impact on performance by modulating achievement motivation, which, in turn, affects effort (Borghans et al., 2008; Segal, 2008). Relatedly, individuals who differ in their personality characteristics also differ with respect to their risk attitude (Becker et al, 2012; Capra et al., 2013), likely by modulating incentive motivation for obtaining large rewards (Capra et al., 2012; Engelmann et al., 2009; Pothos et al., 2011). In the social domain, a link between the Big Five personality traits and trust, as well as reciprocity has been established (Dohmen et al., 2008; Becker et al.,

2012). Specifically, in experimental settings (Dohmen et al., 2008) and large representative samples (Becker et al., 2012), both conscientiousness and emotional reactivity were negatively associated with the propensity to trust. Given that the data on the relationship between personality and social preferences are largely descriptive at this point, the question about the sources of the individual differences in social preferences remains unanswered. The main goal of the present investigation therefore was to probe measures of specific personality characteristics related to anti-social behavioral tendencies to help identify the underlying mechanisms that lead to strategic behavior in social choice settings.

The approach we take here advances prior investigations of the heterogeneity in social preferences in multiple ways: First, we focus on personality-environment interactions to identify individual differences in the strategies people deploy to solve social dilemmas. As implied by standard definitions of personality (Roberts, 2009), the behavioral tendencies assessed in personality questionnaires strongly depend on environmental factors. To give an example, extroversion is typically assessed in social situations ("I enjoy being the center of attention *at a party*"), while emotional reactivity expresses itself in the presence of emotional challenges, such as stress ("When I am very stressed, …"). Clearly, behavior is not consistent across situations (Mischel & Shoda, 1995; Roberts, 2009; Fergusen et al., 2011) and such behavioral plasticity across contexts is an important factor that contributes to human and animal fitness (Dingemanse et al., 2009). Dominant theories incorporate behavioral plasticity into the way they conceptualize personality, such as the cognitive-affective personality systems theory (Mischel & Shoda, 1995) and, more recently, the socio-genomic framework of personality (Roberts, 2009). Cross-situational variability is particularly relevant for social traits, as social environments are inherently complex and dynamic, such that social dilemmas can be solved via multiple approaches that are highly context dependent. Therefore, to identify the strategic nature underlying social decision-making, we manipulate the environment within which social decisions are made in the context of a well-formulated game-theoretic setting, the trust game. Specifically, to investigate the interplay of environmental changes with personality in a standard trust game setting, we allowed the investor to sanction the trustee on a subset of trials. We expected all subjects to adjust their behavior in the punishment environment relative to the sanction-free environment. Importantly, we ask whether specific personality characteristics are associated with an enhanced magnitude of this change in behavior above and beyond a mean change, such that there are specific personality variables that predict a greater adjustment of behavior due to the punishment environment.

A second factor that advances prior investigations of the heterogeneity in social preferences is our focus on measures of anti- and pro-sociality to test hypotheses about strategic interactions in social dilemmas. Prior investigations of the interplay between personality and economic preferences focused almost exclusively on broad dimensions of personality, such as the big five (e.g., Borghans et al., 2008, Beckers et al., 2011). More specific personality traits, such as Machiavellianism, have been relatively understudied (Ferguson et al., 2011). To address this gap in the literature and to test specific hypotheses about the sources of individual differences in social preferences, we focused on personality characteristics that assess pro- and anti-social behavioral tendencies, such as trustworthiness and Machiavellianism. Rather than being merely reactive, anti-social traits are assumed to induce proactive behavior that is characterized by anticipating future outcomes and taking action before situations become a source of confrontation (Jakobwitz & Egan, 2006). Based on these results, we expected strategic, payoff-maximizing choice behavior of people with anti-social characteristics in the context of anonymous one-shot trust game interactions with and without the option to punish low trustee reciprocity. More specifically, payoff-maximizing behavior of the investor (the first mover) would be revealed by an increased investment propensity when given the option to punish and concurrently expectations of higher returns from the trustee. Similarly, payoff-maximizing behavior of the trustee (the second mover) would be revealed by lower reciprocity levels in the absence of punishment and increased reciprocity in its presence, to avoid possible sanctions by the investor within the punishment environment.

Methods

Participants. 182 volunteers (98 females, 84 males; mean age (SD) 22.7 (3.9) years) from various Universities in Zurich participated in the current study. Students with a background in Economics and Psychology were excluded from the experiment to avoid potential decision biases. Participants gave written and informed consent to procedures approved by the local ethics committee (Zurich, Switzerland). No deception was used throughout all procedures employed in the current experiment. Three participants were removed from further analyses due to technical problems (1) and because personality scores identified them as clear outliers (2), leading to a final sample size of N = 179.

Procedure. The experimental procedure was as follows: First, participants filled out an online battery of personality questionnaires some days (mean (SD): 5.0 (3.2) days) before the experiment. Second, they were invited to the UZH laboratory where the experiment was conducted using z-Tree (Fischbacher, 2007). Participants were randomly assigned to the role of either the investor (P1; N = 90) or trustee (P2; N = 89) and never switched roles for the length of the entire experiment. After each trial, all players were randomly matched with anonymous counterparts. In total, there were 13 sessions with 8 to 20 subjects per session, each containing 4 within-subject treatments in counterbalanced order. All treatments were repeated 6 times. Feedback and payout occurred at the end of the session, preventing the emergence of social history and wealth effects. Payoff structure and game setup were common-knowledge. Subjects earned 20 CHF (1 CHF = 1.05 US$) for completing the online questionnaires and received additional payouts for one randomly selected trial from the trust and risk games

Experimental Design. To test specific hypotheses about the influence of personality on social preferences, four different versions of the trust game were administered to all participants in counterbalanced order: the *NPT (no punishment treatment)* and *PT (punishment treatment)* were identical binary trust games, except that investors had the option to punish in the PT. This allowed us to identify the effect of personality on decisions to invest and reciprocate in different punishment environments. Two additional control games were included to ensure the robustness and ecological validity of our results via the *NBT (non-binary treatment)* and to identify the effects of emotional arousal via the *DT (direct feedback treatment)*. In all versions

of the trust game, two players, player 1 (P1), the investor, and player 2 (P2), the trustee, were first endowed with 20 CHF and then sequentially exchanged money as shown in Fig. 1.

Stage 1: P1 moves first and decides in a binary choice situation whether to send nothing to P2 and keep the 20 CHF endowment, in which case the current round ends and both players keep their 20 CHF, or whether to transfer *(T = trust amount)* 10 CHF to P2. If P1 decides to send money, this amount is tripled leading to a wealth distribution of 10 CHF for P1 and 50 CHF for P2 at the beginning of stage 2.

To assess the strategic nature of investment changes due to the option to punish, we also measured the investor's belief about the trustee's back-transfer throughout each version of the trust game. Beliefs of the investor were measured after each investment decision as follows: (1) In all games P1 states his belief about how much he thinks P2 will back-transfer immediately after the transfer decision; (2) in the *PT*, P1 also states his belief about how much he thinks that P2 has transferred back immediately after the decision to punish.

Stage 2: P2 decides how much to transfer back to P1 (*BT = back-transfer amount)* conditional on P1's initial transfer (i.e. 0 CHF or 10 CHF). If P1 chose to invest CHF 10, the choice menu of potential back-transfer amounts ranges from 0 CHF and 50 CHF in 5 CHF increments, if P1 invested nothing, P2's choice menu of potential back-transfer amounts ranges from 0 CHF and 20 CHF. The latter option was included to test for altruistic giving, which occurred on very few occasions. P2 chooses via the strategy method (contingent decisions for all possible transfer amounts from P1; Brandts & Charness, 2011). In each round, the choice is binding, provided that the relevant option has actually been chosen by P1. In the simplest version of the trust game we employed here, the game ends after stage 2, such that final endowments are *(20 – T + BT)* CHF for P1 and *(20 + 3*T – BT)* CHF for P2. The purpose of this treatment was to assess the baseline willingness to trust for P1 and to reciprocate for P2, as it did not include the option to punish.

Stage 3: To identify which personality characteristics were specifically sensitive to unfair behavior in our P1s and the option to be punished in our sample of P2s, we included a version of the trust game in which P1s had the option to punish P2s *(PT).* Specifically, while the first two stages of this version of the trust game were equivalent to the *NPT*, as shown in Fig. 1, we included a third stage, in which P1 chooses via the strategy method what amount *(P = punishment amount)* he or she will spend to reduce P2's payoff. For each CHF invested by P1, 5 CHF will be deducted from P2. This creates an equal balance of power in the *PT*, such that if P2 would back-transfer 0 CHF and P1 spends his remaining 10 CHF to punish P2, both players

would end up with a final payoff of 0 CHF. Note that P2's final payoff is prevented from being negative, even if P1's punishment amount (P * 5) is greater than P2's current holdings.

In our analyses, we focus on the differences in behavior between the *PT* and *NPT*, in order to identify an environment-induced behavioral change and, importantly, whether specific personality characteristics predict an enhanced magnitude of this change in behavior above and beyond a mean change. Results from additional trust games that were employed as robustness checks are reported in SI section S1 for the non-binary game (NBT) and SI section S3 for the direct feedback (DT) game. Of note, the order of three trust games without any feedback information *(NPT, PT, NBT)* was counterbalanced and presentation order was included as a control variable in all regression analyses. Because feedback about the other players' transfer and back-transfer amounts were provided during the *DT*, this game was always presented last to avoid social history and wealth effects.

[Insert Fig. 1]

Risk Tasks. After completion of the trust games, a subset of 104 participants made risky decisions in the context of a certainty equivalent task. The task consisted of a total of 126 individual decisions, in which each choice scenario offered an alternative between choosing a probabilistic lottery and a sure amount. The lottery offered one potential payoff that is greater than the sure amount, and one that is smaller. The payout was determined by randomly selecting 1 of the 126 choice scenarios for which participants earned additional cash amounts between 0 and 50 CHF. For the remainder of subjects (N = 78), risk attitude was assessed via a series of 6 choices between a lottery (same in all trials: 50% chance of winning either 10 CHF or 0.5 CHF) and increasing amounts of safe payments (increasing from 2 CHF to 7.5 CHF). The switch point, reflective of the certainty equivalent in this choice scenario, is taken as a measure of risk attitude. The payout was determined by randomly selecting one of the 6 lotteries for which participants earned additional cash amounts between 0.5 CHF and 10 CHF. Detailed results assessing the relationship between risk and trust, which has been a recurring concern in the literature (Houser, Schunk & Winter, 2010; Altmann, Dohmen & Wibral, 2008) are reported in SI section S4.

Psychological Questionnaires. We assessed personality traits that we hypothesized to have an impact on trust and reciprocity, using well-established psychological questionnaires. Specifically, we included measures of (1) anti- and pro-sociality, (2) impulsivity, (3) emotional reactivity and (4) the big 5 (neuroticism, extraversion, openness, agreeableness and

conscientiousness, assessed via the NEO-FFI; Costa & McCrae, 2006). Measures of *anti- and pro-social behavioral tendencies* included (a) the 27-item PNR scale (Personal Norm of Reciprocity; Perugini et al., 2003) that assesses individuals' propensity for positive and negative reciprocity, as well as subjects' belief in reciprocity; (b) the 25-item MACH-IV questionnaire (Machiavellianism Inventory; Allsopp et al., 1991) that measures strategic, anti-social and selfish behavioral tendencies; (c) the 16-item SDS (Social Desirability Scale; Stöber, 1999) that assesses individual tendencies to act in a socially desirable manner; (d) the Interpersonal Reactivity Index (IRI, Davis, 1983) that assesses various aspects related to empathy, such as perspective-taking ability, the tendency to identify with fictional characters, empathic concern and personal distress; (e) the 28-item PTS questionnaire (Propensity to Trust Scale; Evans & Revelle, 2008) that measures willingness to trust and trustworthiness; (f) the 38-item ECR questionnaire (Experiences in Close Relationships; Ehrenthal et al., 2009) that assesses the tendency to be anxious about and avoid close relationships. Measures of *impulsivity* included (a) the 30-item BIS-11 (Barrat Impulsiveness Scale; Patton, Stanford & Barrat, 1995) that assesses "attentional", "motor" and "non-planning" impulsivity; and (b) the 40-item SSSV (Sensation-Seeking Scale V; Zuckermann, 1994) that measures aspects of sensation seeking, such as thrill and adventure seeking, disinhibition, experience seeking and boredom susceptibility. Measures of emotional reactivity included (a) the 24-item STAXI (Intensity and Disposition to Experience State and Trait Anger; Spielberger, 2010) that measures the intensity and disposition to experience anger; (b) the 20-item STAI (a measure of state and trait anxiety; Spielberger, 2010) that assesses anxiety; (c) the 21-item BDI scale (Beck Depression Inventory; Beck et al., 1996) that measures depression severity; (d) the 24-item BIS/BAS (Behavioral Inhibition/Activation System; Carver & White, 1994) that measures the tendency to be motivated by appetitive and aversive behavioral outcomes. Finally, immediately after the behavioral tasks, subjects' current mood state and perceived stress were assessed by the 12-item MDBF questionnaire (Steyer et al., 1997) that measures current mood, arousal and anxiety state, and the 10-item PSS (Perceived Stress Scale; Cohen, Kamarck, & Mermelstein, 1983) that measures the perception of stress and the degree to which situations in one's current life are appraised as stressful.

Results

Factor Analysis. Given the high dimensionality of our personality data, we first employed factor analysis to reduce the data to the most essential elements and remove sources of

covariance before entering this data into regression analysis. This method has three distinct advantages for the current investigation: (1) It removes potential redundancies by identifying the essential elements in the questionnaire data that are important for further analysis of the influence of personality on decision making in changing environments; (2) it creates a set of uncorrelated factors that explain the unique contribution of specific personality factors to social decision-making, and (3) it reduces noise due to measurement inaccuracies by producing new variables from redundant items. An exploratory factor analysis using maximum likelihood estimation with orthogonal varimax rotation in Stata 14.2 was conducted for the 37 questionnaire items* and 5 factors were retained, indicated jointly by the scree-plot and the latent-root-criterion. Factor loadings reported in table 1 can be interpreted as correlations between factors and the corresponding item, with a higher loading making the item more representative of the factor.

[Insert Table 1]

Table 1 shows a well interpretable and sensible factor structure suggesting the following classification of the retained 5 factors: **(1)** The emotional reactivity *(EMO) factor* shows high positive loadings on subscales reflective of high levels of emotional reactivity, such as trait anxiety, depression and neuroticism, as well as anger suppression and anxiety about close relationships and negative loadings on extroversion and trust; **(2)** the anti-sociality *(ANTI) factor* shows high positive loadings on anti-social subscales reflective of Machiavellianism, financial and ethical risk taking (DOSPERT) and avoidance of relationships, as well as high negative loadings on pro-social subscales reflective of trustworthiness, empathic concern, and agreeableness; **(3)** the sensation seeking *(SS) factor* shows high positive loadings on all subscales related to sensation seeking (SSS-V), such as disinhibition, boredom susceptibility, experience seeking and especially thrill and adventure seeking, as well as health and recreational risk taking (DOSPERT), **(4)** the anger *(ANG) factor* shows high positive loadings on subscales reflective of trait anger, aggressive behavior and negative reciprocity, as well as high negative loadings on anger control and social desirability; **(5)** the impulsivity *(IMP) factor* shows high positive loadings on all measures of impulsivity (BIS-11), such as attentional and motor, but especially non-planning impulsiveness and high negative loadings on conscientiousness, reward responsiveness and goal oriented behavior. To construct the five

* Items measuring state variables, such as state anger, state anxiety, perceived stress and current mood were not included in the factor analysis. The following subscales were not included as they exhibit low KMO (Kaiser-Meyer-Olkin measure of sampling adequacy) and lead to decreasing Cronbach's alpha (internal consistency) measures: NEO-FFI openness, BAS fun-seeking and PNR belief.

compound personality variables for subsequent analyses, factor scores for the latent variables were computed using Bartlett's approach, producing normalized and unbiased maximum likelihood estimates of the "true" factor scores for each participant (e.g. Costello & Osborne, 2005; Hayton, Allen & Scarpello, 2004).

Our main research goal was to assess the extent to which personality factors predict the influence of punishment on (1) trust, as reflected by investor decisions to transfer money to another anonymous player, (2) reciprocity, as reflected by the trustee's decision to reciprocate, and (3) punishment propensity, as reflected in the investor's decision to invest money in reducing the trustee's final payout. We addressed these research questions by investigating the association between personality factors and choice behavior at each stage of the trust game, by first analyzing (1) P1's trust, then (2) P2's reciprocity and finally, (3) P1's punishment behavior.

Stage 1: Trust. We performed multiple Logit regressions estimating the association between personality and trust in the presence and absence of punishment to address the following questions: What is the general influence of personality on the investor's decision to trust? Does the option to punish trust betrayal influence the investor's propensity to trust? To what degree does personality affect the magnitude of the impact of the option to punish on trust? We modeled the influence of punishment on trust via the dummy variable *PT* (punishment present (1), absent (0)). In addition, our models included a number of socio-economic control variables[†], namely *Sex* (male/female), *City* (reflective of living in a city with > 10'000 inhabitants), *Swiss* (reflective of cultural background) and *Age*, as well as controls for mood *(MDBF)* and stress level *(PSS)* that were measured at the time of the experiment. Our results are not dependent on these a priori confounders, as all results hold without these control variables[‡]. To account for the panel structure (within-subject design in which each individual played six rounds in each treatment)[§], cluster-robust standard errors were calculated.
Table 2 presents the estimated coefficients for three different models: Model 1 only contains socio-economic control variables, as well as *PSS* and *MDBF* scores and the treatment dummy *PT*. Model 2 adds the five personality variables, and model 3 additionally its interactions with

[†] Note: Game-specific control variables, such as treatment-order dummies, round-one dummy and session size were included in all our models but are not reported in the tables.
[‡] In most cases, effects of confounders are not significantly different from zero, thus are not reported in the tables.
[§] As proposed by Barr (2013) when testing interaction effects with multiple observations per unit, mixed-effects models with random intercept and slope (for the within-subject treatment effect *PT*) coefficients were tested for both game stages, but qualitative results did not change.

the treatment environment, which are of particular importance to the current research question as interactions address whether personality predicts the magnitude of the effect that the option to punish has on trust beyond its mean effect. The following discussion will focus on the results from model 3, because this model yielded the *lowest* corrected *AIC* value (*AICc;* Wagenmakers & Farrell, 2004) and our research questions focus on the interaction terms.

[Insert Table 2, Fig. 2]

On average, P1s armed with the option to punish P2s exhibited a 15 percentage points increase in their probability to transfer 10 CHF ($p < 0.01$; Mean transfer probability PT: 74.3%; NPT: 60.7%). Conversely, higher ANG factor scores led to a significantly lower transfer probability (if *ANG factor* scores would increase by 1 unit, this probability decreases by 8 percentage points). The same magnitude was observed for the *ANTI factor*, however showing only marginal significance ($p < 0.1$). Notably, an interaction between anti-sociality and the option to punish was observed. As shown in Fig. 2A, this interaction indicates that the relationship between the propensity to trust and anti-sociality changes in the presence compared to the absence of punishment. Specifically, participants with high *ANTI factor* scores demonstrated a significantly increased transfer probability in the punishment relative to the no-punishment environment ($p < 0.01$). As reported in Table 2 and illustrated in Fig. 2, no other personality factor (apart from anger, which also showed a significant, but much less substantial marginal probability effect) significantly interacts with the punishment environment.

Such shifts in behavior are likely of a strategic nature, in that the punishment option may lead to enhanced back-transfer expectations in subjects with higher anti-sociality, but not in others. Increasing "trust" under such beliefs would be strategically optimal as it is expected to generate greater payouts. To test this notion, we assessed (1) whether specifically anti-sociality and no other personality factor is associated with greater earnings in the trust game, and (2) the change in investors' beliefs about how much they think their counterparts will back-transfer in the presence compared to the absence of punishment. We find that anti-sociality is significantly and positively associated with higher potential earnings. Specifically, we find an average increase in potential earnings of 2.10 CHF in the absence of punishment ($p < 0.001$) and of 1.25 CHF in the presence of punishment ($p < 0.05$). This positive association was not observed for any other personality factor (NPT: all $p > 0.246$, PT: all $p > 0.083$).

Beliefs about backtransfers are shown in Table 3 and illustrated in Fig. 3. These results demonstrate a significant and positive association between expectations of back-transfer amounts and anti-sociality in the presence relative to the absence of punishment ($p < 0.01$). This result supports the hypothesis that investors with greater anti-sociality act more strategically, as they adjust their investment behavior based on their increased back-transfer expectations when given the option to punish subsequent digressions. Similar results were obtained for the ANG factor, except that anger did not significantly interact with the punishment treatment. Given that no other personality factor showed this relationship, the strategic expectation-based nature of investment changes due to the option to punish is therefore specific to anti-social personality characteristics.

[Insert Table 3]

Stage 2: Trustworthiness. We performed multiple OLS regression[**] estimating the association between personality and back-transfers in the presence and absence of punishment (similar procedure as in stage 1) to address the following questions: What is the influence of personality on the trustee's decision to reciprocate and does the option to punish influence the trustee's reciprocation? To what degree does personality affect the magnitude of the punishment effect on reciprocity? First, we identified an association between anti-sociality and back-transfer amounts, such that greater anti-sociality was associated with lower average back-transfer amounts ($p < 0.01$, this effect was reproduced when considering back-transfers in the non-binary setting, see Fig. S1B and Table S2). Next, we focused on the impact of the investors' option to punish trustees (modeled via the dummy variable *PT*) on the association between back-transfer and personality. Table 4 shows that the possibility of being punished by the investor increased trustee back-transfers by 3.50 CHF on average (mean back-transfer PT: 14.81 CHF, *NPT*: 11.30 CHF, $p_{delta} < 0.01$). Furthermore, higher anti-sociality *(ANTI)* and anger *(ANG) factor* scores led to a significantly lower back-transfer on average (ANTI: $p < 0.01$; ANG: $p < 0.05$). Of note, a significant interaction between anti-sociality and the presence of punishment was observed, indicating that the degree of a person's anti-sociality affects the magnitude of the change in reciprocal behavior in the presence compared to the absence of the

[**] We used the strategy method to record responses from trustees, obtaining back-transfer rates for two potential scenarios, namely for the case in which the investor sent 0 CHF and 10 CHF. The former was included to test pro-social motives, such as altruism, which we did not observe (an average of a mere 0.5 CHF was returned when P1 sent 0 CHF). We therefore focused on the case in which P1's initial transfer had been 10 CHF.

investor's option to punish ($p < 0.05$). Specifically, the greater the anti-sociality, the greater was the discrepancy between relatively low back-transfers when P1 could not punish and relatively higher back-transfers when P1 was able to do so (Fig. 4). Such shifts in back-transfer amounts in the punishment environment are likely of a strategic nature, such that subjects with higher anti-sociality scores may anticipate greater emotional discontent and subsequent punishment from investors who face low back-transfers. Indeed, as we show below, anti-social characteristics are associated with significantly stricter punishment of digressions (Fig. 5A). Under conditions of heightened expectations to be punished, sending back higher monetary amounts is payoff maximizing. One common mechanism to form expectations about the behavior of counterparts in social interactions is self-projection (e.g., Silani et al., 2013), particularly in the absence of relevant information about other players, as is the case in the anonymous one-shot trust games played in the current experiment. We therefore investigated the relationship between anti-sociality and punishment as a proxy for the expectation to get punished. Specifically, demonstrating that anti-social characteristics are associated with greater punishment of deviations from expectations is consistent with the notion that anti-social personality characteristics lead to greater back-transfer amounts via an accurate self-projection mechanism.

[Insert Table 4, Fig. 4]

Stage 3: Punishment. To what degree does the willingness to punish depend on personality? Furthermore, is the relationship between back-transfer amount and costly punishment magnitude affected by personality? To answer these questions, we first assessed the average relationship between back-transfer amounts and punishment, given that P1's initial transfer had been 10 CHF[††]. To capture the non-linear, negative exponential relationship between back-transfer and punishment amounts, we adapted models typically employed to explain intertemporal choice behavior (Kable & Glimcher, 2007; Laibson, 1997; see SI Equations S8-S12). This specification allows the slope parameters of the exponential function to change for different back-transfer domains, as further discussed below. Model comparison revealed that a quasi-hyperbolic, double-exponential functional model best fits the data (Equation S11; Figure S2; for model comparison see Table S3). We found the expected relationship between back-transfer amount and punishment magnitude, such that subjects punish very low back-transfers more severely than back-transfers around a point of perceived fairness (at 25 CHF as estimated

[††] P1s that have transferred 0 CHF in the first stage did not expect positive back-transfer amounts and also did not spend substantial amounts to punish P2s, independent of back-transfer amounts.

via piecewise OLS; Equation S7; Fig. S2). Specifically, as reported in Table 5, there is a decline in average punishment until the point of perceived fairness, reflected by the negative slope coefficients BT1 and BT2 (both $p < 0.05$), and little average punishment after this point (this effect is further pronounced by BT2 being less negative than BT1; $p_{Delta} < 0.01$[‡‡]). All models were estimated via non-linear least squares regression. Model comparison was conducted via Akaike weights based on corrected AIC (Table S3; Wagenmakers & Farrell, 2004).

Next, we answered the question whether personality predicts the severity of punishment (Table 5 and Fig. 5). We found that subjects with high anti-social characteristics exhibited the strongest punishment response to trustee (P2) defection, as indicated by a significant positive effect of anti-sociality on the intercept ($p < 0.01$), followed by a strong average decline in punishment amounts as back-transfers increased, especially in the low back-transfer domain, as indicated by the negative slope effect (ANTI x BT2; $p < 0.05$) of anti-sociality interacted with back-transfer. Specifically, anti-sociality was associated with significant increases in punishment of low back-transfers: For zero back-transfers, a unit increase in the *ANTI factor* score led to an increase in punishment amount by 1.98 CHF on average (see Equation S13 for the marginal effect of the *ANTI factor*). This gap became smaller with higher back-transfers and finally disappeared after a back-transfer of about 25 CHF. Furthermore, impulsivity was also associated with significantly higher punishment of defection, but showing only half of the magnitude compared to anti-sociality (marginal effect of the *IMP factor* at zero back-transfer = 0.95 CHF; $p < 0.05$). Finally, a higher *EMO factor* score diminished the response function slope, leading to a reduced punishments of low back-transfer amounts ($p < 0.05$).

[Insert Table 5: Fig. 5]

‡‡ Slope coefficients in this model have to be interpreted *relative* to the slope coefficient BT1 in the exponential model; see Equation S8 and Kable & Glimcher, 2007

Discussion

The main goal of the current investigation was to assess the role of specific personality traits in explaining heterogeneity in social preferences. Specifically, we hypothesized that anti- and pro-social characteristics interact with environmental challenges in the production of social behavior. Through independent self-report measures, we first identified a personality profile that clearly reflects anti-social personality characteristics, with high positive loadings on Machiavellianism and high negative loadings on empathy, trustworthiness and agreeableness. When looking at the role of personality in the context of a standard trust game without punishment, anti-sociality predicts decision strategies in a manner that is consistent with its name: significantly lower levels of trust and decreased trustworthiness.

To identify the strategic nature of anti-social behavior in changing environments, we assessed the interaction between personality and environmental changes on investor and trustee decision-making in the context of trust games with and without punishment. Simply the knowledge that the investor can sanction the trustee leads to significant increases in investor trust and trustee reciprocity across all subjects. Anti-social personality characteristics predict the magnitude of this environmental effect above and beyond a mean effect. Specifically, greater anti-sociality is associated with significant increases in trust taking and reciprocity within the punishment environment. No other personality factor was predictive of such a strong adjustment of decision strategies in the presence of punishment. Taken together, these results demonstrate that anti-social personality characteristics are associated with strategic behavioral shifts aimed at maximizing the extraction of resources from their counterparts in the context of social dilemmas. This strategy was successful, as anti-sociality, but no other personality factor, was associated with greater earnings.

Our results also inform the mechanisms underlying the shifts in choice strategy associated with anti-sociality. Specifically, we show that anti-sociality is associated with significant changes of investors' beliefs in changing environments. Importantly, such shift in beliefs across punishment conditions was specific for anti-sociality and was not observed for any other personality factors. These results indicate that one mechanism through which anti-sociality affects decision strategies is through a belief structure that is significantly more context-dependent compared to other personality factors.

The question arises what psychological mechanism underlies the context-dependent decision strategies and belief structure of anti-social investors and trustees. The surprising finding that exclusively anti-sociality is associated with specific changes in decision strategies in the

punishment context across all three stages of the trust game implies the following: (1) the significant increase in trust that is specific to anti-social investors reflects an accurate prediction about the increase in back-transfer amounts under conditions of punishment; (2) this is paralleled by the significantly more optimistic expectations about trustee back-transfer amounts under conditions of punishment, particularly for anti-social investors; and (3) trustees with high levels of anti-sociality accurately predict the severe punishment of low back-transfers by investors with high levels of anti-sociality. The significant changes in trust and beliefs across punishment environments displayed by anti-social investors and trustees are therefore optimally tailored for interactions with other anti-social types as they accurately predict the behaviour of their anti-social interaction partners. Together, these observations are consistent with the notion that decision strategies and beliefs of persons with high trait levels of anti-sociality are based on self-projection, which is the process of simulating one's own thoughts, feelings and intentions within a hypothetical scenario to make predictions about the behavior of others (Waytz & Mitchell, 2011). To give an example, to estimate the likelihood of betrayal (Bohnet and Zeckhauser, 2004; Aimone et al., 2014) investors might simulate how they would act if they were in the role of the trustee, while trustees might simulate how they would react to low back-transfer amounts to estimate the likelihood of retaliation. In the context of anonymous one-shot trust games in which no further identifying information about counterparts is provided, self-projection may be an important and adaptive process, especially if the motivation of the player is to maximize payoffs.

The behavioral profile outlined above agrees with predictions made about anti-social personality characteristics in the clinical literature. Anti-social types show a good understanding of the intentions and emotions of others by anticipating behavior under different settings (Whiten & Byrne, 1997), while at the same time showing deficits in empathizing abilities in social interrelations (Lyons, Caldwell & Schultz, 2010). Self-projection is one plausible mechanism that can enable such exploitative strategies in the competition for resources observed in anti-social types, as it can lead to egocentricity bias (e.g., Silani et al., 2013) and lack of empathy, particularly in the case of a person with payoff maximizing motives. In the current study, we were able to predict such strategic behavioral tendencies in the context of a trust game from assessments using standard personality questionnaires in a psychologically normal student population.

Our results provide an important empirical step in developing an economic theory of anti-social behavior. First, they show that specific personality factors can modulate choice strategies across different environments. Specifically, anti-sociality is associated with a flexible behavioral

response to different economic environments. Our evidence suggests that such behavioral plasticity is employed by persons with high levels of anti-sociality whose motive it is to maximize payoffs by extracting resources from other players. No other personality profile showed similar levels of behavioral plasticity in the context of the social interactions captured by the trust games employed in the current study, indicating the specificity of context-dependent decision strategies for anti-sociality.

Finally, it is worth mentioning that our research approach provides a number of methodological advances that may be important for future research at the intersection between Personality Psychology and Behavioral Economics. As suggested by dominant theories of personality and captured by a recent economic framework (Almlund et al., 2012), behaviors can change across situational contexts. Context gains additional relevance when it comes to social decision strategies, as social interactions are inherently complex and highly context-dependent. Therefore, clearly conceptualized environmental manipulations are important for understanding individual differences in choice strategies and to identify the role of personality in decision-making in general. Our results therefore suggest that future research should adapt the approach taken in the current investigation and manipulate contexts relevant for the choice strategies under investigation, as otherwise important associations between personality and economic behavior across situations may not be identifiable.

Acknowledgments. We gratefully acknowledge financial support from the NCCR Affective Sciences. We thank Carola Hug for assistance with data collection and Joel van der Weele and Isabel Thielmann for comments on an earlier version of this manuscript.

References

Aimone, J.A., Houser, D., Weber, B. (2014). Neural signatures of betrayal aversion: an fMRI study of trust. Processdings of the Royal Society B, 281(1782), 20132127

Allsopp, J., Eysenck, H. J. & Eysenck, S. B. G. (1991). Machiavellianism as a Component in Psychoticism and Extraversion. *Person. Individ. Diff.*, 12(1), 29-41.

Almlund, M., Duckworth, A. L., Heckman, J. J. & Kautz, T. (2011). Personality Psychology and Economics. *Discussion Paper*, (5500).

Altmann, S., Dohmen, T. & Wibral, M. (2007). Do the Reciprocal Trust Less? *Economic Letters*, 99, 454-457.

Antonakis, J., Bendahan, S., Jacquart, P. & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21(6), 1086-1120.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4(328), 1-2.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment,* 67(3), 588-597.

Becker, A., Falk, A., Deckers, T., Kosse, F. & Dohmen, T. (2012). The Relationship between Economic Preferences and Psychological Personality Measures. *Discussion Paper,* (6470).

Berg, J., Dickhaut, J.& McCabe, K. (1995). Trust, Reciprocity and Social History. *Games and Economic Behavior*, 10(1), 122-142.

Bohnet, I. & Zeckhauser, R. (2004) Trust, risk and betrayal. Journal of Economic Behavior & Organization, 55, 467-484.

Boksem, M. A. S., Tops, M., Meijman, T. F. & Lorist, M. M. (2006). Error-Related ERP Components and Individual Differences in Punishment and Reward Sensitivity. *Brain Research*, 1101, 92-101.

Borghans, L., Duckworth, A. L., Heckman, J. J. & Weel, B. (2008). The Economics and Psychology of Personality Traits. *Discussion Paper*, (3333).

Bowles, S. & Gintis, H. (2004). Social Norms and Human Cooperation. *Theoretical Population Biology*, 65, 185-190.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. (2003). The Evolution of Altruistic Punishment. *Proceedings of the National Academy of Sciences*, 100(6), 3531-3535.

Brandts, J. & Charness, G. (2011). The Strategy Versus the Direct-Response Method: A First Survey of Experimental Comparisons. *Experimental Economics*, 14(3), 375-398.

Burks, S. V., Carpenter, P. J., & Verhoogen, E. (2003). Playing Both Roles in the Trust Game. Journal of Economic Behavior & Organization, 51, 195-216. Capra, M. (2004). Mood-Driven Behavior in Strategic Interactions. *The American Economic Review*, 94(2).

Capra, M., Jiang, B., Engelmann, J. B. & Berns, G. (2013). Can Personality Type Explain Heterogenety in Probability Distortions? *Journal of Neuroscience, Psychology, and Economics*, 6(3), 151-166.

Carver, Ch. S. & White, T. L. (1994). Behavioral Inhibition, Behavioral Activation and Affective Responses to Impending Reward and Punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology,* 67(2), 319-333.

Cohen, M. X., Schoene-Bake, J. Ch., Elger, Ch. E. & Weber, B. (2010). Connectivity-Based Segregation of the Human Striatum Predicts Personality Characteristics. *Nature Neuroscience*, 1-3.

Cohen, S.,Karmarck, T. & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, 24(4), 385-396.

Costa, P. T. & McCrae, R. R. (2006). The Five Factor Theory of Personality. In O. P. John, R.W. Robins & L. A. Persvin (Eds.). Handbook of Personality: Theory and Research, The Guilford Press.

Costello, A. B. & Osborne, J. W (2005). Best Practices in Explanatory Factor Analysis: Four Recommendations for Getting the Most From Your Data. *Practical Assessment Research & Evaluation*, 10(7), 1-9.

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G. & Robbins, T. W. (2010). Impulsive Choice and Altruistic Punishment Are Correlated and Increase in Tandem with Serotonin Depletion. *Emotion*, 10(6), 855-862.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Journal of Human Resources, 43, 738-782.

Cunha, F., Heckman, J. J., & Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78, 883–931.

Davis, M.H. (1983) Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of Personality and Social Psychology,* 44(1), 113-126

Dingemanse, N. J., Kazem, A. J. N., Reale, D., & Wright, J. (2010). Behavioral reaction norms: Animal personality meets individual plasticity. *Trends in Ecology and Evolution*, 25, 81–89.

Dohmen, T., Falk, A., Huffman, D. & Sunde, U. (2008). Representative Trust and Reciprocity: Prevalence and Determinants. *Economic Inquiry,* 46(1), 84-90.

Ehrenthal, J. C., Dinger, U., Lamla, A., Funken, B., & Schauenburg, H. (2009). Evaluation of the German Version of the Attachment Questionnaire ”Experiences in Close Relationships - Revised” (ECR-RD). *Psychother Psych Med,* 59, 215-223.

Elster, J. (1989). On the Economics and Biology of Trust. *Journal of the European Economic Association*, 3(4), 99-117.

Engelmann, J.B. (2006). Personality Predicts Responsivity of the Brain Reward System. *The Journal of Neuroscience*, 26(30), 7775-7776.

Engelmann, J.B., Damaraju, E., Padmala, S. & Pessoa, L. (2009). Combined Effects of Attention and Motivation on Visual Task Performance: Transient and Sustained Motivational Effects. *Frontiers in Human Neuroscience*, 3(4).

Eshel, N. & Roiser, J. P. (2010). Reward and Punishment Processing in Depression. *Biol Psychiatry*, 68, 118-114.

Evans, A. M. & Revelle, W. (2008). Survey and Behavioral Measurements of Interpersonal Trust. *Journal of Research in Personality*, 42, 1585-1593.

Fehr, E. (2009). On the Economics and Biology of Trust. *Journal of the European Economic Association*, 7(2-3), 235-266.

Fehr, E., & Fischbacher, U. (2004). Social Norms and Human Cooperation. *Trends in Cognitive Sciences*, 8(4), 185-190.

Fehr, E., Fischbacher, U. & Gächter, S. (2002). Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms. *Human Nature*, 13(1), 1-25

Fehr, E., & Gächter, S. (2002). Altruistic Punishment in Humans. *Nature*, 415, 137-140.

Ferguson, E., Heckmann, J. J. & Corr, Ph. (2011). Personality and Economics: Overview and Proposed Framework. *Personality and Individual Differences*, 51, 201-209.

Fischbacher, U. (2007). z-tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, 10(2), 171-178.

Fischbacher, U., Gächter, S. & Fehr, E. (2001). Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. *Economic Letters*, 71, 307-404.

Fowler, J. H. (2005). Altruistic Punishment and the Origin of Cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, 12(9), 7047-7049.

Gunnthorsdottir, A., McCabe, K. & Smith, V. (2002). Using the Machiavellianism Instrument to Predict Trustworthiness in a Bargaining Game. *Journal of Economic Psychology*, 23, 49-66.

Hayton, J. C., Allen, D. G. & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods*, 7(2), 195-205.

Houser, D., Schunk, D. & Winter, J. (2010). Distinguishing Trust from Risk: An Anatomy of the Investment Game. *Journal of Economic Behavior & Organization*, 74, 72-81.

Imai, K., Keele, L., Tingley, D. & Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*, 105(4), 765-789.

Jakobwitz, S. & Egan, V. (2006). The Dark Triad and Normal Personality Traits. *Personality and Individual Differences*, 40, 331–339.

Jones, D. N. & Paulhus, D. L. (2010). Different Provocations Trigger Aggression in Narcissists and Psychopaths. *Social Psychological and Personality Science,* 1(1), 12-18.

Kable, J. W. & Glimcher, P. W. (2007). The Neural Correlates of Subjective Value during Intertemporal Choice. *Nature Neuroscience*, 10(12), 1625-1633.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2), 443-478.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The Influence of Probability on Risky Choice: A Parametric Examination. *Journal of Economic Behavior & Organization*, 17(3), 377-400.

Lyons, M., Caldwell, T. & Schultz, S. (2010). Mind-Reading and Manipulation – Is Machiavellianism Related to Theory of Mind?. *Journal of Evolutionary Psychology*, 8(3), 261-274.

McCabe, K., Rigdon, M. & Smith, V. (2003). Positive Reciprocity and Intentions in Trust Games. *Journal of Economic Behavior & Organization*, 52, 267-275.

Meyer, H. D. (1992). Norms and Self-Interest in Ultimatum Bargaining: The Prince's Prudence. *Journal of Economic Psychology*, 13, 215-232.

Mischel, W., Shoda, Y., 1995. A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. Psychol. Rev. 102 (2), 246-268.

Must, A., Szabo, Z., Bodi, N., Szasz, A., Janka, Z. & Keri, S. (2006). Sensitivity to Reward and Punishment and the Prefrontal Cortex in Major Depression. *Journal of Affective Disorders*, 90, 209-215.

Norton, E. C., Wang, H. & Ai, Ch. (2004). Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal*, 4(2), 154-167.

Oud, B., Williams, T., Engelmann, J. B., Krumhuber, E., & Fehr, E. (2012). Facial Cues and Trust-Related Behavior. *Working Paper.*

Patton, J. H., Stanford, M. S. & Barrat, E. S. (1995). Factor Structure of the Barrat Impulsiveness Scale. *Journal of Clinical Psychology*, 51(6), 768-774.

Perguini, M., Gallucci, M., Presaghi, F. & Ercolani, A. P. (2003). The Personal Norm of Reciprocity. *European Journal of Peronality*, 17(4), 251-283.

Ridderinkhof, K. R., Wildenberg, W. P. M., Sidney, J. S. & Carter, C. S. (2004). Neurocognitive Mechanisms of Cognitive Control: The Role of Prefrontal Cortex in Action Selection, Response Inhibition, Performance Monitoring, and Reward-Based Learning. *Brain and Cognition*, 56, 129-140.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. Journal of Research in Personality, 43, 137-145.

Segal, C. (2008). Motivation, test scores, and economic success. Department of Economics and Business, Universitat Pompeu Fabra, Working Paper No. 1124.

Spielberger, C. (2010). State-Trait Anxiety Inventory. Corsini Encyclopedia of Psychology.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF). Handanweisung, Göttingen.

Stöber, J. (1999). The Social Desirability Scale - 17 (SDS-17: Development and First Findings on Reliability and Validity). *Diagnostica*, 45(4).

Wagenmakers, E. J. & Farrell, S. (2004). AIC Model Selection using Akaike Weights. Psychonomic Bulletin & *Review*, 11(1), 192-196.

Weber, E. U., Blais, A. R. & Betz, N. E. (2002). A Domain-Specific Riskattitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making,* 15, 263-290.

Whiten, A. & Byrne, R. (1997). Machiavellian Intelligence II. Cambridge University Press.

Zuckerman, M. (1994). Behavioral Expressions and Biosocial Bases of Sensation Seeking. Cambridge University Press.

Table 1: Results from an exploratory factor analysis of 37 personality questionnaire items included in the online questionnaire. Five factors can be identified that are reflective of emotional reactivity (EMO factor), anti-sociality (ANTI factor), sensation seeking (SS factor), trait anger (ANG factor), and impulsivity (IMP factor). The factor analysis was conducted by maximum likelihood estimation with orthogonal varimax rotation (Kaiser normalization). Kaiser-Meyer-Olkin measure of sampling adequacy: 0.82. Cronbach's Alpha: 0.83. LR test: 5 factors vs. saturated: Prob. < 0.00. Variance explained: 48.8 %. Number of Subjects: 179. Subject-to-Item ratio: 4.8. Blanks: |loading| < 0.35. Bold: |loading| > 0.5.

Item	EMO Factor	ANTI Factor	SS Factor	ANG Factor	IMP Factor	Uniqueness
PTS TRUST	**-0.5588**	-0.4567				0.4584
PTS TRUSTWT		**-0.5778**		-0.3931		0.3202
SDS				**-0.5283**		0.5072
MACH IV		**0.6192**				0.4869
IRI PT		-0.3939				0.6584
IRI FS		-0.3786				0.7961
IRI EC		**-0.7730**				0.3939
IRI PD	**0.6113**					0.5432
STAI T	**0.8959**					0.1580
STAXI T	0.3826			**0.6664**		0.3723
STAXI AO				**0.6797**		0.5210
STAXI AC				**-0.5637**		0.6214
STAXI AI	**0.5889**					0.6170
BIS BI	**0.6612**					0.3824
BAS DRIVE					**-0.6119**	0.5554
BAS REWARD		-0.4048			-0.4219	0.6204
BIS 11 AT	0.3667				0.3928	0.5584
BIS 11 MT			0.4777		0.4437	0.5325
BIS 11 NP					**0.6291**	0.4604
DOSPERT ET		0.4662		0.3818		0.5047
DOSPERT FI		0.4768				0.6948
DOSPERT HE			**0.5738**			0.6177
DOSPERT RE			**0.8652**			0.2128
DOSPERT SP			0.4074			0.7991
PNR POS		-0.4502				0.7257
PNR NEG				0.4868		0.5845
NEO NR	**0.8402**					0.2246
NEO EX	**-0.5432**	-0.3907				0.4884
NEO AG		**-0.6948**		-0.4049		0.3144
NEO CN					**-0.7151**	0.3663
ECR BANXIETY	**0.5878**					0.6280
ECR BAVOIDANCE		**0.5326**				0.6686
BDI	**0.6974**					0.4237
SSSV TA			**0.8256**			0.2047
SSSV DI			**0.5184**			0.6364
SSSV EX			**0.5373**			0.6463
SSSV BD		0.3589	0.4437			0.6359

Table 2: Logit regressions investigating the effect of personality on investor decisions to trust (*T*) in the presence compared to the absence of the option to punish. The dependent variable in all columns is the investor's decision to trust. *PT* is a dummy variable reflective of the option to punish. To assess the influence of personality on trust, all five factors were included in the model, including emotional reactivity (EMO), anti-sociality (ANTI), sensation seeking (SS), trait anger (ANG), and impulsivity (IMP). Differential effects of personality in the presence compared to the absence of punishment were modeled by interacting all personality factors with the dummy variable reflective of the presence of punishment (factor x *PT*). Constant, treatment-order dummies, round-one dummy, session size, sex, city, Swiss, age, MDBF and PSS are not reported in the table. Results from model 3 are reported in the main paper. Marginal interaction effects were derived by calculating the discrete differences with respect to the dummy variable *PT* of the single derivatives with respect to the continuous factor variables (Norton, Wang & Ai, 2004). All covariates were fixed at their means. Significance levels: *** = 1%, ** = 5% and * = 10%.

Model	1	2	3	3
Variable	Coef. / (SE)	Coef. / (SE)	Coef. / (SE)	dy/dx / (SE)
PT	0.665***	0.696***	0.716***	0.15***
	(0.15)	(0.16)	(0.16)	(0.03)
EMO Factor	-	0.141	0.190	0.04
		(0.21)	(0.23)	(0.05)
ANTI Factor	-	-0.120	-0.379*	-0.08*
		(0.20)	(0.22)	(0.05)
SS Factor	-	0.267	0.205	0.04
		(0.17)	(0.18)	(0.04)
ANG Factor	-	-0.267	-0.389**	-0.08**
		(0.17)	(0.18)	(0.04)
IMP Factor	-	0.145	0.058	0.02
		(0.17)	(0.18)	(0.04)
EMO x PT	-	-	-0.083	-0.03
			(0.16)	(0.04)
ANTI x PT	-	-	0.546***	0.12***
			(0.19)	(0.04)
SS x PT	-	-	0.178	0.02
			(0.13)	(0.03)
ANG x PT	-	-	0.236	0.06**
			(0.15)	(0.03)
IMP x PT	-	-	0.148	0.02
			(0.18)	(0.04)
N (# Clusters)	1080 (90)	1080 (90)	1080 (90)	
AICc	1300	1264	1257	
Pseudo R^2	0.07	0.10	0.12	
Prob. > X^2	0.00	0.00	0.00	

Table 3: OLS regressions investigating the effect of personality on investors' beliefs in the presence compared to the absence of the option to punish. The dependent variable in all columns is P1's belief about P2's back transfer. *PT* is a dummy variable reflective of the option to punish. Differential effects of personality in the presence compared to the absence of punishment were modeled by interacting all personality factors with the dummy variable reflective of the presence of punishment (factor x PT). Constant, treatment-order dummies, round-one dummy, session size, sex, city, Swiss, age, MDBF and PSS are not reported in the table. Significance levels: *** = 1%, ** = 5% and * = 10%.

Model	1	2	3
Variable	Coef. / (SE)	Coef. / (SE)	Coef. / (SE)
PT	2.676***	2.676***	2.634***
	(0.60)	(0.52)	(0.52)
EMO Factor	-	0.041	0.636
		(0.63)	(0.67)
ANTI Factor	-	-0.882	-1.913**
		(0.80)	(0.76)
SS Factor	-	0.954	0.721
		(0.59)	(0.64)
ANG Factor	-	-1.631***	-2.022***
		(0.54)	(0.59)
IMP Factor	-	0.358	0.419
		(0.67)	(0.67)
EMO x PT	-	-	-1.191**
			(0.57)
ANTI x PT	-	-	2.062***
			(0.60)
SS x PT	-	-	0.466
			(0.49)
ANG x PT	-	-	0.783
			(0.60)
IMP x PT	-	-	0.121
			(0.60)
N (# Clusters)	1080 (90)	1080 (90)	1080 (90)
AICc	7751	7700	7687
R^2	0.12	0.17	0.18
Prob. > X^2	0.00	0.00	0.00

Table 4: OLS regressions investigating the effect of personality on trustee back transfer (*BT*) decisions in the presence compared to the absence of the option to punish. The dependent variable in all columns is the trustee's back transfer amount when the investor transferred 10 CHF. *PT* is a dummy variable reflective of the option to punish. To assess the influence of personality on reciprocity, all five factors were included in the model, including emotional reactivity (EMO), anti-sociality (ANTI), sensation seeking (SS), trait anger (ANG), and impulsivity (IMP). Differential effects of personality in the presence compared to the absence of punishment were modeled by interacting all personality factors with the dummy variable reflective of the presence of punishment (factor x PT). Constant, treatment-order dummies, round-one dummy, session size, sex, city, Swiss, age, MDBF and PSS are not reported in the table. Results from model 3 are reported in the main paper. Constant, treatment-order dummies, round-one dummy and session size are not reported in the table. Results from model 3 are reported in the main paper. Significance levels: *** = 1%, ** = 5% and * = 10%.

Model	1	2	3
Variable	Coef. / (SE)	Coef. / (SE)	Coef. / (SE)
PT	3.511***	3.511***	3.458***
	(0.78)	(0.79)	(0.76)
EMO Factor	-	0.454	0.822
		(0.85)	(0.89)
ANTI Factor	-	-1.772**	-2.584***
		(0.68)	(0.74)
SS Factor	-	0.491	0.321
		(0.68)	(0.80)
ANG Factor	-	-1.493**	-1.659**
		(0.62)	(0.71)
IMP Factor	-	0.612	1.205*
		(0.49)	(0.65)
EMO x PT	-	-	-0.736
			(0.71)
ANTI x PT	-	-	1.623**
			(0.63)
SS x PT	-	-	0.340
			(0.79)
ANG x PT	-	-	0.331
			(0.65)
IMP x PT	-	-	-1.187
			(0.73)
N (# Clusters)	1068 (89)	1068 (89)	1068 (89)
AICc	7568	7480	7465
R^2	0.10	0.18	0.20
Prob. > *F*	0.00	0.00	0.00

Table 5: Non-linear regressions investigating the effect of personality on investor punishment behavior. The dependent variable in all columns is the amount invested to punish the trustee when the investor transferred 10 CHF. Results from the three best models as identified by model comparison (see Table S3) are shown. The top three models are shown here, all of which have an exponential functional form. BT1 and BT2 are slope parameters reflective of a reduction in average punishment amounts invested as back-transfers increased. For the exponential model, there was one slope parameter BT1, whereas for the quasi-hyperbolic models (Kable & Glimcher, 2007) there were two: Given the more negative value of BT1 in the two quasi-hyperbolic models, BT1 reflects a special weight placed on lower back-transfers, whereas BT2 is reflective of higher back-transfers. Therefore, Table 5 indicates that there is a change in estimated slope parameters, showing a more curved shape (compared to the exponential model) when back-transfers were lower. To assess the influence of personality on punishment, all five factors were included in the model, including emotional reactivity (EMO), anti-sociality (ANTI), sensation seeking (SS), trait anger (ANG), and impulsivity (IMP). Constant, treatment-order dummies, round-one dummy, session size, sex, city, Swiss, age, MDBF and PSS are not reported in the table. Results from model 3 are discussed in the main paper. Significance levels: *** = 1%, ** = 5% and * = 10%.

Model	Exponential	Quasi-hyp. red.	Quasi-hyp. full
Variable	Coef. / (SE)	Coef. / (SE)	Coef. / (SE)
BT1	-0.072***	-0.122***	-0.095**
	(0.01)	(0.04)	(0.04)
BT2	-	-0.055***	-0.070**
		(0.01)	(0.03)
EMO Factor	0.027	-0.047	0.052
	(0.48)	(0.45)	(0.46)
ANTI Factor	1.770***	1.909***	1.979***
	(0.48)	(0.47)	(0.46)
SS Factor	-0.128	-0.092	-0.140
	(0.42)	(0.41)	(0.40)
ANG Factor	0.406	0.327	0.500
	(0.43)	(0.37)	(0.38)
IMP Factor	0.766*	0.777*	0.953**
	(0.45)	(0.46)	(0.47)
EMO x BT1	0.017**	0.056***	0.023**
	(0.01)	(0.02)	(0.01)
ANTI x BT1	0.004	-0.044***	0.025
	(0.01)	(0.02)	(0.02)
SS x BT1	0.001	0.006	0.004
	(0.01)	(0.02)	(0.01)
ANG x BT1	0.005	0.003	0.014
	(0.01)	(0.01)	(0.01)
IMP x BT1	0.011*	0.023*	0.017
	(0.01)	(0.01)	(0.01)
EMO x BT2	-	-	0.004
			(0.01)
ANTI x BT2	-	-	-0.043**
			(0.02)
SS x BT2	-	-	-0.012
			(0.01)
ANG x BT2	-	-	-0.020
			(0.01)
IMP x BT2	-	-	-0.012
			(0.01)
N (# Clusters)	4481 (80)	4481 (80)	4481 (80)
AICc	18305	18203	18073
R^2	0.41	0.42	0.44
Prob. > X^2	0.00	0.00	0.00

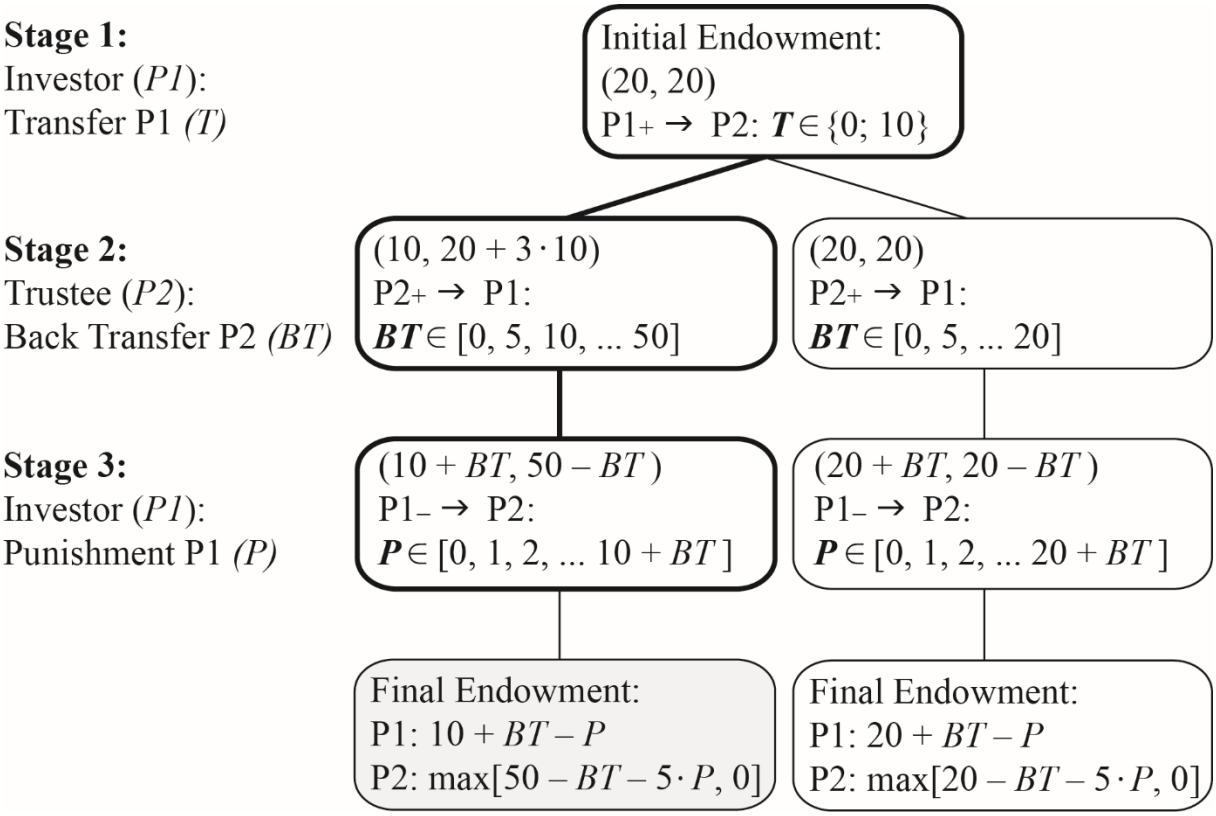

Fig. 1: Schematic representations of the trust game setup for the two main treatments, the *NPT* (no punishment treatment; the game ends after stage 2) and the *PT* (punishment treatment; the game ends after stage 3). All values are in Swiss Francs (1 CHF = 1.05 US$). **Left:** Game-tree where P1's initial transfer is 10 CHF. **Right:** Game-tree where P1's initial transfer is 0 CHF. We obtained back transfer rates for two potential scenarios, namely for the case in which the investor sent 0 CHF and 10 CHF. The former was included to test pro-social motives, which we did not observe (an average of a mere 0.5 CHF was returned when P1 sent 0 CHF). All analyses therefore focus on the case in which P1's initial transfer had been 10CHF.

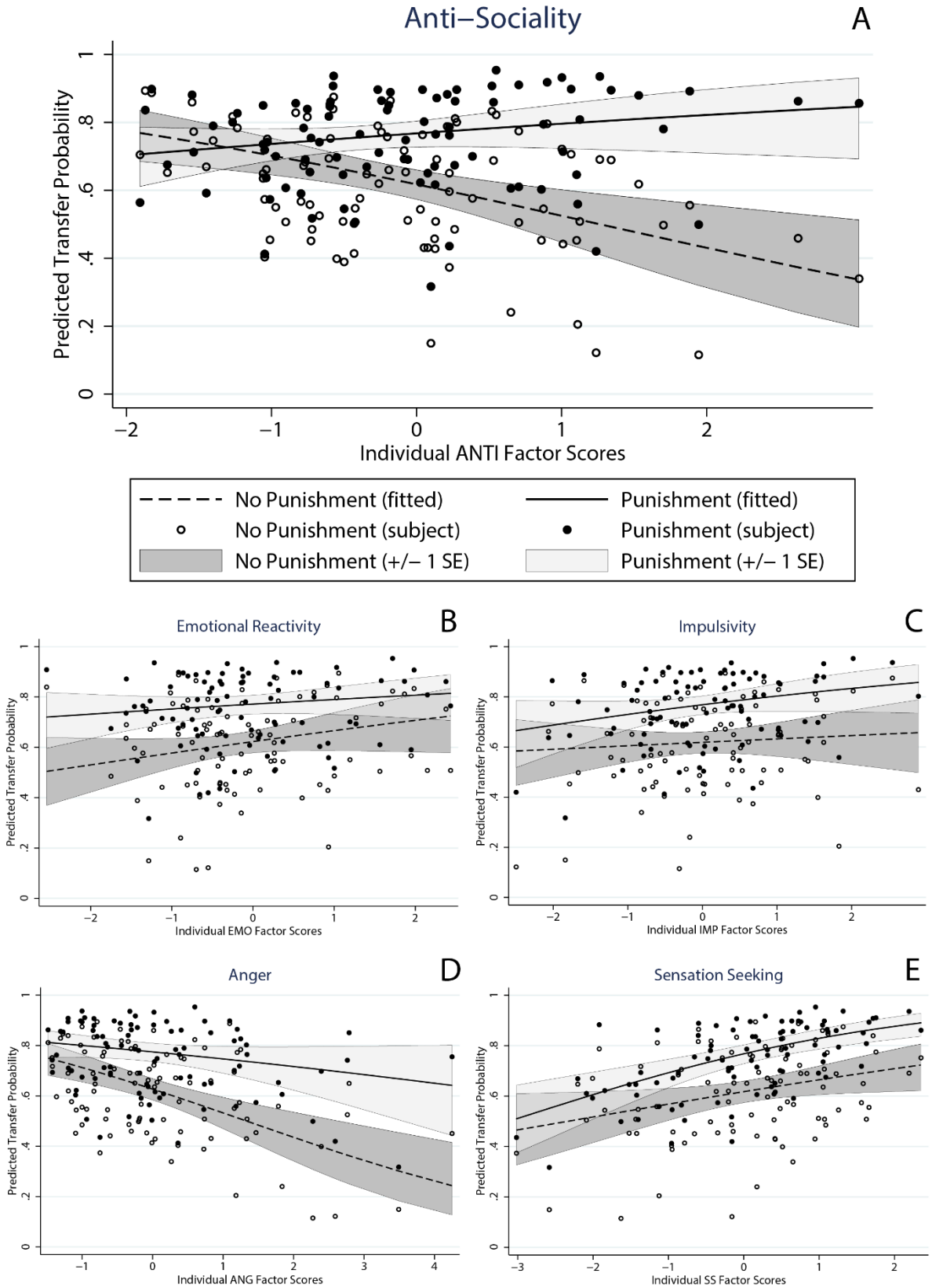

Fig. 2: Investor transfer amounts as a function of punishment condition (absent vs. present) and personality factor scores. Graphs visualize main, treatment and interaction effects for each of the five personality traits (A: ANTI factor; B: EMO factor; C: IMP factor; D: ANG factor; E: SS factor). Predictions are based on model 3 (Table 2). **Dots:** P1's predicted transfer probabilities for each subject (N = 90) in the no punishment treatment (*NPT*) and the punishment treatment (*PT*). **Solid/dashed lines:** Fitted values. All covariates (except factor, treatment and interaction of interest) were fixed at their means.

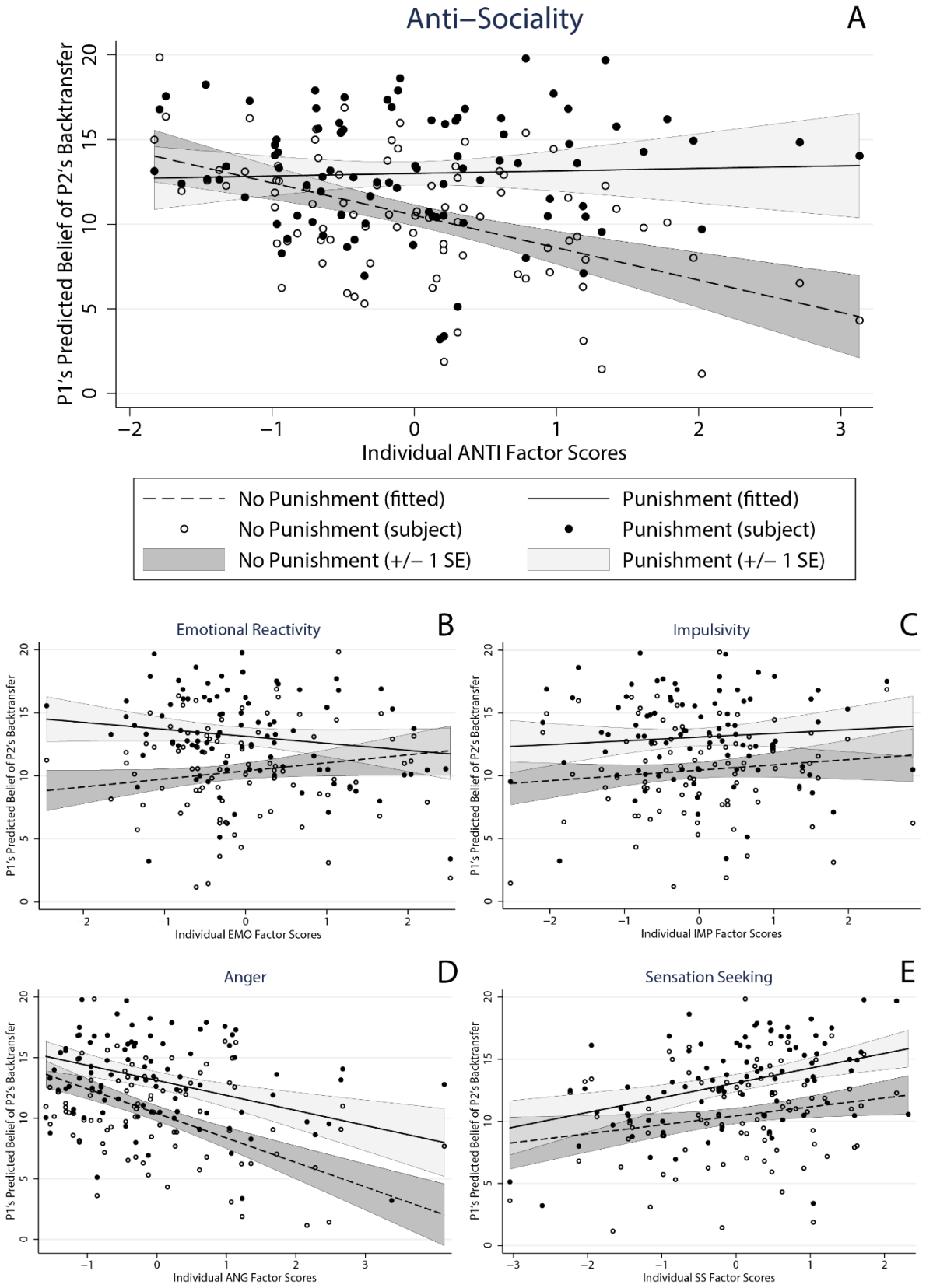

Fig. 3: Investor belief about trustee reciprocity as a function of punishment condition (absent vs. present) and personality factor scores. Graphs visualize main, treatment and interaction effects for each of the five personality traits (A: ANTI factor; B: EMO factor; C: IMP factor; D: ANG factor; E: SS factor). Predictions are based on model 3 (Table 3). **Dots**: P1's predicted belief about P2's back transfer for each subject (N = 90) in the no punishment treatment (NPT) and the punishment treatment (PT). **Solid/dashed lines**: Fitted values. All covariates (except factor, treatment and interaction of interest) were fixed at their means.

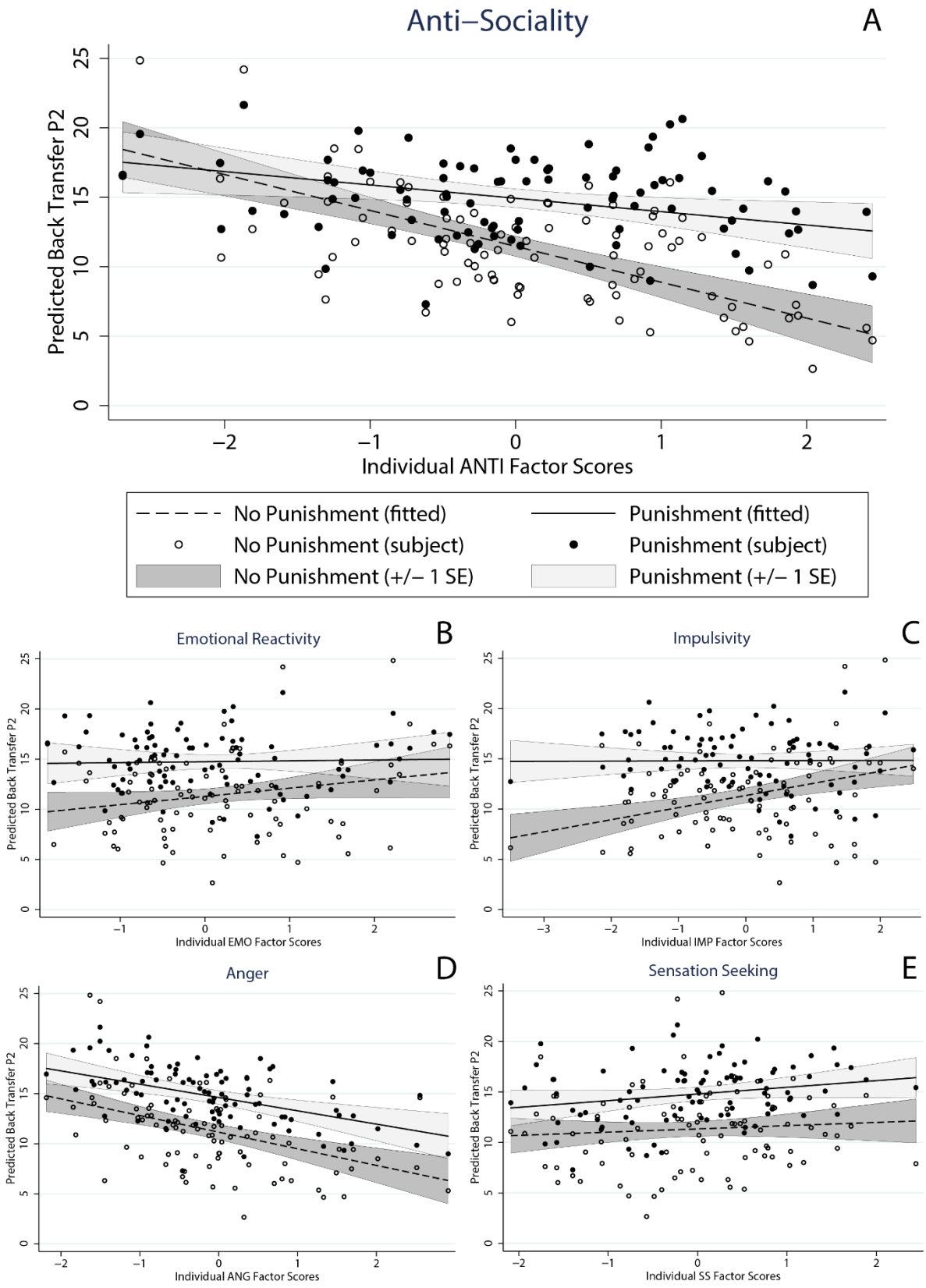

Fig. 4: Trustee back-transfer amounts as a function of punishment condition (absent vs. present) and personality factor score. Graphs visualize main, treatment and interaction effects for each of the five personality traits (A: ANTI factor; B: EMO factor; C: IMP factor; D: ANG factor; E: SS factor). Predictions are based on model 3 (Table 4). **Dots:** P2's average back transfers for each subject (N = 89) in the no punishment treatment (*NPT*) and the punishment treatment (*PT*). **Solid/dashed lines:** Fitted values. All covariates (except factor, treatment and interaction of interest) were fixed at their means.

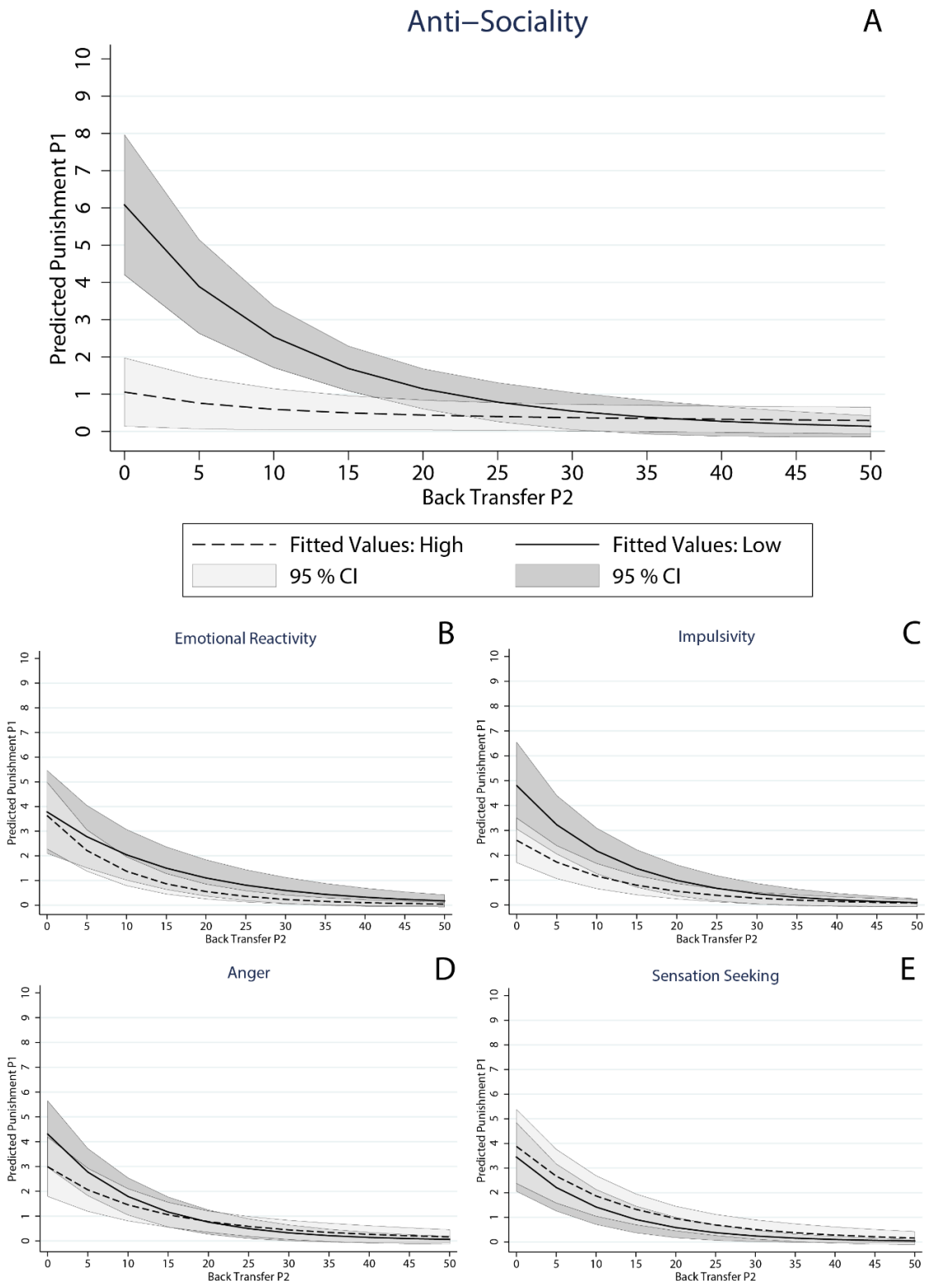

Fig. 5: Investor punishment amounts for given trustee reciprocity and each of the five personality traits (A: ANTI factor; B: EMO factor; C: IMP factor; D: ANG factor; E: SS factor), given investor's initial transfer was 10 CHF (N = 80). Predictions are based on model 3 (Table 5; Equation S11). All covariates were fixed at their means. The factors of interest ("high" vs. "low") were fixed at the 90th and 10th factor score percentiles, respectively. Note: Average factor score differences between the 90th and 10th percentiles varied between 2.5 and 3.1 factor score units (*ANTI factor* score difference btw. 90th – 10th p.c. = 2.78 factor score units). The estimated marginal effect of a unit increase in *ANTI factor* scores on punishment severity, *given zero back transfers,* was 1.98 CHF ($p < 0.01$).

Supporting Information for "The Dark Side of Personality: Anti-Sociality Increases Strategic Game Play"

Engelmann, J. B., Schmid, B., Chumbley, J. and Fehr, E. (2018)

S.1 Non-binary treatment (NBT)

Trustee back-transfer amounts exhibited a non-linear increase for increasing investor transfer amounts in the non-binary version of the trust game (NBT). In order to capture these non-linearities and identify the influences of personality factors on trustees' reciprocity, we compared five classes of possible back-transfer response functions: A piecewise linear model, an exponential model, a standard hyperbolic model, a quasi-hyperbolic model according to Laibson (1997) and a quasi-hyperbolic model according to Kable & Glimcher (2007). The goal was to assess differences in effects of personality (i.e. factors) on back-transfers, including parameters estimating personality related differences in back-transfer response function slopes (i.e. factors interacted with investor transfer amounts).

Table S1: Akaike weights: Stage 2 model selection. **Model S1:** Piecewise linear model. The kink-point was estimated at T = 9 CHF. Therefore, model S1 includes a dummy variable (= 1) for investor transfer amounts $\geq$ 9 CHF, $D9_{i,r,bt}$, and its interaction with transfer, allowing the intercept and slope to change after the kink. **Model S2:** Exponential model. **Model S3:** Standard hyperbolic model. **Model S4:** Quasi-hyperbolic model according to Laibson (1997). **Model S6:** Quasi-hyperbolic, reduced model according to Kable & Glimcher (2007): Transfer x factor interactions were only included once.

Model	# Parameters	AICc	Δ_i	Akaike w_i
S1: Piecewise linear	27	31061	240	0
S2: Exponential	25	30831	10	0.0077
S3: Standard hyperbolic	25	31194	373	0
S4: Standard quasi-hyperbolic	25	30833	12	0
S6: Quasi-hyperbolic (reduced)	26	30821	0	0.9923

Model selection was conducted using Akaike weights, w_i, based on small sample size corrected AIC values, $AICc$ (Wagenmakers & Farrell, 2004; Burnham & Anderson, 2002). Akaike weights reflect the probability that a given model, M_i, is the best model among the alternative models, given the data. Table S1 presents a comparison between five different model specifications. The following model specifications were employed to investigate the effect of personality on trustee reciprocity:

Piecewise linear:

$$BT_{i,r,t} = c + \alpha F_i + \lambda T_{i,r,t} + \beta[T_{i,r,t} \cdot F_i] + \rho D9_{i,r,t} \tag{S1}$$
$$+\kappa[D9_{i,r,t} \cdot T_{i,r,t}]$$

Exponential:

$$BT_{i,r,t} = (c + \alpha F_i) \cdot \exp(\lambda T_{i,r,t} + \beta[T_{i,r,t} \cdot F_i]) \tag{S2}$$

Standard hyperbolic:

$$BT_{i,r,t} = (c + \alpha F_i) \cdot \frac{1}{1 + \lambda T_{i,r,t} + \beta[T_{i,r,t} \cdot F_i]} \tag{S3}$$

Standard quasi-hyperbolic:

$$BT_{i,r,t} = (c + \alpha F_i) \cdot (\lambda + \beta F_i])^{T_{i,r,t}} \tag{S4}$$

Quasi-hyperbolic (full):

$$BT_{i,r,t} = (c + \alpha F_i) \cdot \frac{1}{2}\{\exp(\lambda T_{i,r,t} + \beta[T_{i,r,t} \cdot F_i]) + \tag{S5}$$
$$+\exp(\omega T_{i,r,t} + \tau[T_{i,r,t} \cdot F_i])\}$$

Quasi-hyperbolic (reduced):

$$BT_{i,r,t} = (c + \alpha F_i) \cdot \frac{1}{2}\{\exp(\lambda T_{i,r,t} + \beta[T_{i,r,t} \cdot F_i]) + \tag{S6}$$
$$+\exp(\omega T_{i,r,t})\}$$

- $BT_{i,r,t}$: Back-transfer for each individual i, round r and transfer t
- F_i : Vector containing all five factors for each individual i
- c : Constant, treatment-order-dummies, round-one-dummy, session size and other control variables
- $T_{i,r,t}$: Vector containing all ten possible transfers (i.e. 1 - 10 CHF in steps of 1 CHF)
- λ : Slope coefficient ($T1$)
- ω : Slope coefficient ($T2$)
- α : Five coefficients representing constant level effects of personality factors (F_i)
- β : Five coefficients representing slope effects of personality factors w.r.t. transfers ($F_i \cdot T1$)
- τ : Five coefficients representing slope effects of personality factors w.r.t. transfers ($F_i \cdot T2$)

The three winning models are presented in Table S2. Model S6 is the winner (see also Fig. S1 A), speaking in favor of a quasi-hyperbolic back-transfer re-

sponse pattern w.r.t. initial transfers (of note, the full quasi-hyperbolic model, in which personality factors were interacted with both curvature parameters, could not be estimated due to collinearity issues). Importantly, all models yielded the same qualitative results irrespective of model specification:
(1) There was a significant non-linear increase in the amount that was back-transferred as investor transfers increased, as indicated by significant slope parameter $T2$;
(2) Anti-sociality had a strong effect on average back-transfer amounts ($p < 0.01$, Fig. S1 B), indicative of a general reduction of trustee back-transfers no matter what the investor initially transferred;
(3) Sensation seeking led to a positive effect on back-transfers via the slope parameter $T1$ ($p < 0.05$, Fig. S1 C), indicative of a faster reduction of back-transfer amounts with decreasing investor transfers.

Of note, model S6 includes two slope parameters, one being reflective of trustee back-transfers to low investor transfers ($T1$) that was not significantly different from zero, and $T2$ being reflective of higher trustee back-transfers to increasing investor transfers ($p < 0.01$; given that the coefficient $T2$ is larger than $T1$, the former reflects a special weight placed on higher transfers; $T1$ and $T2$ in this model have to be interpreted relative to $T1$ in the exponential model). Fig. S1 B and Fig. S1 C illustrate the effects (predicted back-transfers in model S6) for "low" and "high" (10th vs. 90th factor score percentiles) anti-social and sensation seeking trustees, respectively.

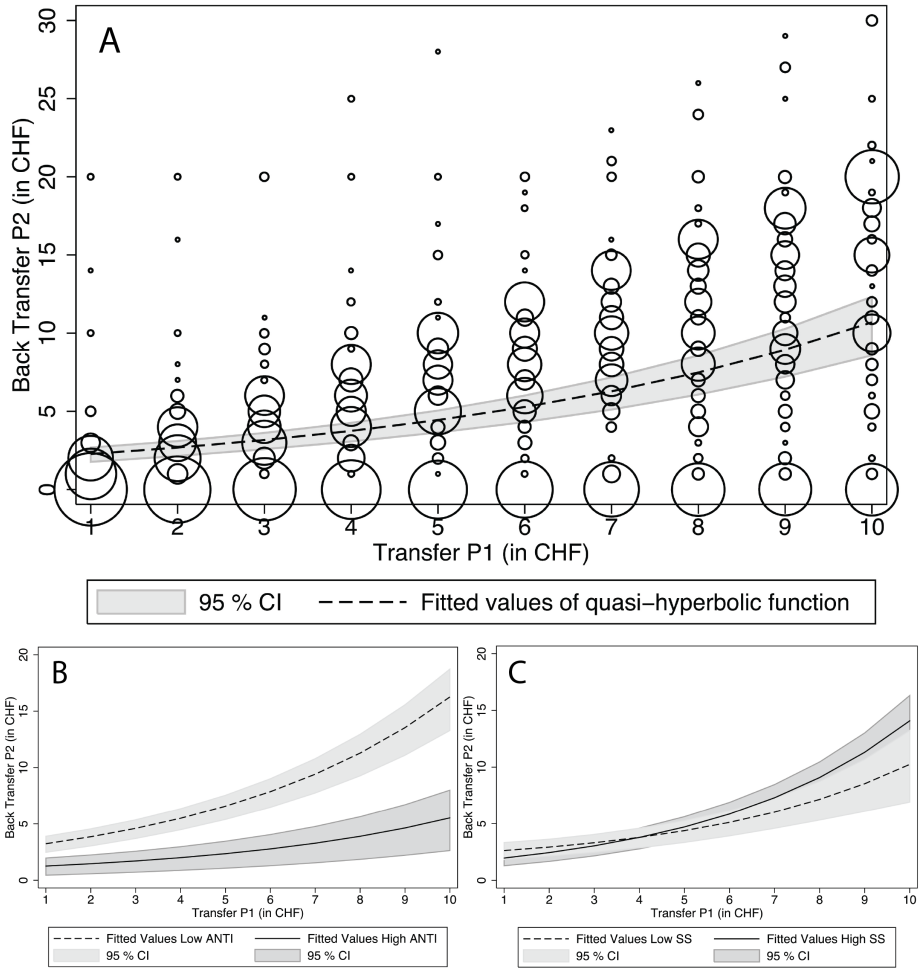

Fig. S1: **A:** Back-transfer amounts of P2, given P1's initial transfer. Note: Circle areas are proportional to the frequency of duplicated observations in each point of the Transfer/BackTransfer-space. Fitted values (model S6): All covariates were fixed at their means. **B:** Quasi-hyperbolic back-transfer response functions for "low" and "high" anti-social trustees (model S6). Fitted values: Values of ANTI Factor were fixed at $10th$ and $90th$ percentile, respectively. All other covariates were fixed at their means. **C:** Quasi-hyperbolic back-transfer response functions for "low" and "high" sensation seeking trustees (model S6). Fitted values: Values of SS Factor were fixed at $10th$ and $90th$ percentile, respectively. All other covariates were fixed at their means.

Table S2: Non-linear regressions investigating the effect of personality on the trustees' back-transfers (BT) in the non-binary treatment without punishment (NBT). The model included a number of control variables which are gender (male = 1), age, as well as culture encoded as being Swiss or not, and city, reflective of living in a city with $> 10'000$ inhabitants. Results from the three best models as identified by model comparison (see Table S1) are presented. Constant, treatment-order-dummies, round-one-dummy and session size are not reported in the table.

Model	S4	S2	S6
Variable	Coef./(SE)	Coef./(SE)	Coef./(SE)
T1	1.209***	0.189***	0.117
	(0.01)	(0.01)	(0.07)
T2	–	–	0.205***
			(0.02)
Male	0.315	0.325	0.411*
	(0.22)	(0.22)	(0.24)
City	0.207	0.205	0.133
	(0.22)	(0.22)	(0.23)
Swiss	-0.229	-0.229	-0.246
	(0.22)	(0.22)	(0.25)
Age	0.086***	0.087***	0.101***
	(0.03)	(0.03)	(0.03)
PSS	0.004	0.005	0.017
	(0.03)	(0.03)	(0.03)
MDBF	0.008	0.008	0.019
	(0.04)	(0.04)	(0.04)
EMO Factor	0.123	0.140	0.295
	(0.17)	(0.17)	(0.19)
ANTI Factor	-0.685***	-0.673***	-0.565***
	(0.17)	(0.17)	(0.18)
SS Factor	-0.364**	-0.362**	-0.286
	(0.16)	(0.16)	(0.17)
ANG Factor	-0.149	-0.162	-0.216
	(0.22)	(0.23)	(0.22)
IMP Factor	0.092	0.109	0.304
	(0.16)	(0.16)	(0.19)
EMO x T1	-0.001	-0.002	-0.026
	(0.01)	(0.01)	(0.02)
ANTI x T1	0.009	0.006	-0.019
	(0.01)	(0.01)	(0.02)
SS x T1	0.044***	0.037***	0.091**
	(0.01)	(0.01)	(0.04)
ANG x T1	-0.003	-0.002	-0.016
	(0.02)	(0.01)	(0.03)
IMP x T1	0.007	0.005	-0.016
	(0.01)	(0.01)	(0.02)
N (# Clusters)		5340 (89)	
R^2	0.52	0.52	0.52
Prob. $> \chi^2$	0.00	0.00	0.00

Significance levels: *** = 1%, ** = 5% and * = 10%.
Robust standard errors, clustered by subjects.

S.2 Punishment response functions

Investor punishment amounts exhibited a non-linear increase for decreasing trustee back-transfer amounts. In order to capture these nonlinearities and identify the influences of personality factors on punishment behavior, we compared five classes of possible punishment response functions: A piecewise linear model, an exponential model, a standard hyperbolic model, a quasi-hyperbolic model according to Laibson (1997) and a quasi-hyperbolic model according to Kable & Glimcher (2007). The goal was to assess differences in effects of personality (i.e. factors) on the amount spent for punishment, including parameters estimating personality related differences in punishment response function slopes (i.e. factors interacted with back-transfer).

Table S3: Akaike weights: Stage 3 model selection. **Model S7:** Piecewise linear model. The kink-point was estimated at back-transfer = 25 CHF. Therefore, model S7 includes a dummy variable (= 1) for back-transfers $\geq$ 25 CHF, $D25_{i,r,bt}$, and its interaction with back-transfer, allowing the intercept and slope to change after the kink. **Model S8:** Exponential model. **Model S9:** Standard hyperbolic model. **Model S10:** Quasi-hyperbolic model according to Laibson (1997). **Model S11:** Quasi-hyperbolic, full model according to Kable & Glimcher (2007): Double-exponential framework, where the different slope parameters (λ vs. ω for the average back-transfer response, and β vs. τ for the back-transfer x factor interactions) reflect a special weight placed on either lower or higher (depending on their relative size) back-transfers. **Model S12:** Quasi-hyperbolic, reduced model according to Kable & Glimcher (2007): Back-transfer x factor interactions were only included once.

Model	# Parameters	AICc	Δ_i	Akaike w_i
S7: Piecewise linear	27	18544	471	0
S8: Exponential	25	18305	232	0
S9: Standard hyperbolic	25	18461	388	0
S10: Standard quasi-hyperbolic	25	18308	235	0
S11: Quasi-hyperbolic (full)	31	18073	0	1
S12: Quasi-hyperbolic (reduced)	26	18203	130	0

Model selection was conducted using Akaike weights, w_i, based on small sample size corrected AIC values, $AICc$ (Wagenmakers & Farrell, 2004; Burnham & Anderson, 2002). Akaike weights reflect the probability that a given model, M_i, is the best model among the alternative models, given the data. Table S3 presents a comparison between six different model specifications. The following model specifications were employed to investigate the effect of personality on investor punishment behavior:

Piecewise linear:

$$P_{i,r,bt} = c + \alpha F_i + \lambda BT_{i,r,bt} + \beta[BT_{i,r,bt} \cdot F_i] + \rho D25_{i,r,bt} \quad \text{(S7)}$$
$$+\kappa[D25_{i,r,bt} \cdot BT_{i,r,bt}]$$

Exponential:

$$P_{i,r,bt} = (c + \alpha F_i) \cdot \exp(\lambda BT_{i,r,bt} + \beta[BT_{i,r,bt} \cdot F_i]) \tag{S8}$$

Standard hyperbolic:

$$P_{i,r,bt} = (c + \alpha F_i) \cdot \frac{1}{1 + \lambda BT_{i,r,bt} + \beta[BT_{i,r,bt} \cdot F_i]} \tag{S9}$$

Standard quasi-hyperbolic:

$$P_{i,r,bt} = (c + \alpha F_i) \cdot (\lambda + \beta F_i])^{BT_{i,r,bt}} \tag{S10}$$

Quasi-hyperbolic (full):

$$P_{i,r,bt} = (c + \alpha F_i) \cdot \frac{1}{2}\{\exp(\lambda BT_{i,r,bt} + \beta[BT_{i,r,bt} \cdot F_i]) + \tag{S11}$$
$$+\exp(\omega BT_{i,r,bt} + \tau[BT_{i,r,bt} \cdot F_i])\}$$

Quasi-hyperbolic (reduced):

$$P_{i,r,bt} = (c + \alpha F_i) \cdot \frac{1}{2}\{\exp(\lambda BT_{i,r,bt} + \beta[BT_{i,r,bt} \cdot F_i]) + \tag{S12}$$
$$+\exp(\omega BT_{i,r,bt})\}$$

- $P_{i,r,bt}$: Amount spent for punishment of each individual i, round r and back-transfer bt
- F_i : Vector containing all five factors for each individual i
- c : Constant, treatment-order-dummies, round-one-dummy, session size and other control variables
- $BT_{i,r,bt}$: Vector containing all eleven possible back-transfers (i.e. 0 - 50 CHF in steps of 5 CHF)
- λ : Slope coefficient ($BT1$)
- ω : Slope coefficient ($BT2$)
- α : Five coefficients representing constant level effects of personality factors (F_i)
- β : Five coefficients representing slope effects of personality factors w.r.t. back-transfers ($F_i \cdot BT1$)
- τ : Five coefficients representing slope effects of personality factors w.r.t. back-transfers ($F_i \cdot BT2$)

The three winning models and a description of the results are presented in the main paper (Table 5). Model S11 was the clear winner (see also Fig. S2), speaking in favor of a quasi-hyperbolic punishment response pattern w.r.t. back-transfers. Given that the slope coefficient $BT1$ is more negative than $BT2$, the former reflects a special weight placed on lower back-transfers (both p $<$

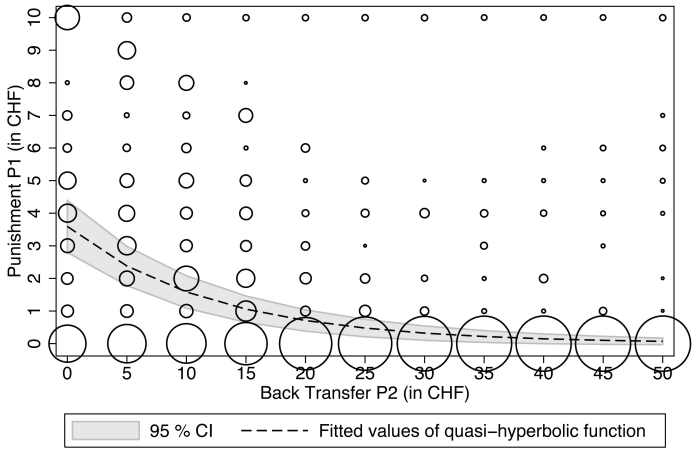

Fig. S2: Amount spent by P1s to punish P2s conditional on P2s' back-transfers (P), given T = 10 CHF. Note: Circle areas are proportional to the frequency of duplicated observations in each point of the BackTransfer/Punishment-space. Fitted values (model S11): All covariates were fixed at their means.

0.05; $BT1$ and $BT2$ in this model have to be interpreted relative to $BT1$ in the exponential model). Note, however, that the main story did not change, no matter what kind of model specification we applied:
(1) Investor punishment amounts decreased non-linearly with increasing trustee back-transfer amounts, reaching near zero after back-transfers of 25 CHF or higher;
(2) Anti-sociality exhibited a strong positive effect ($p < 0.01$) on the amount spent for punishing low back-transfers, but decreasing rapidly for higher back-transfers;
(3) Impulsivity exhibited a positive level effect ($p < 0.05$) on punishment;
(4) Emotional reactivity flattened out the slope of the punishment response function ($p < 0.05$), especially in the low back-transfer domain.

Equation S13 shows the partial derivative of the punishment response function (model S11) with respect to the ANTI factor (f_2). For a back-transfer (BT) of 0 CHF, a unit increase in ANTI factor scores led to an increase in punishment of α_2 = 1.98 CHF. This effect became smaller for larger back-transfers ($P_{i,r,bt}$ is strictly monotonically decreasing in BT) and finally disappeared.

$$\frac{\partial P_{i,r,bt}}{\partial f_2} = (c + \alpha F_i) \cdot BT \cdot \frac{1}{2}\{\beta_2 \cdot \exp(\lambda BT_{i,r,bt} + \beta[BT_{i,r,bt} \cdot F_i]) + \tag{S13}$$

$$\tau_2 \cdot \exp(\omega BT_{i,r,bt} + \tau[BT_{i,r,bt} \cdot F_i])\} + \alpha_2 \cdot \frac{1}{2}\{\exp(\lambda BT_{i,r,bt} +$$

$$\beta[BT_{i,r,bt} \cdot F_i]) + \exp(\omega BT_{i,r,bt} + \tau[BT_{i,r,bt} \cdot F_i])\}$$

S.3 Strategy method (PT) vs. direct feedback (DT)

We investigated the influence of personality on differential trust taking and reciprocity behavior in direct feedback and strategy method settings. We find a significant level effect that reflects enhanced trust taking and enhanced back-transfer amounts in the direct feedback environment DT (on average, 14 %-points higher transfer probability and 3 CHF higher back-transfer; $p < 0.01$).

We find no differential effect of personality on the propensity to trust in the arousing direct feedback relative to the no feedback environment. For trustee reciprocity, we replicate two results reported in the analyses in Table 4, namely that both anti-sociality and anger lead to marginally significant decreases in back-transfer rates ($p < 0.1$). Furthermore, we observe a marginally significant interaction between anti-sociality and the environment, indicating that anti-sociality is associated with increased back-transfer rates in the direct feedback environment relative to the strategy method ($p < 0.1$, Table S4). This result indicates that immediate feedback about punishment enhances the strategic shift in back-transfer amounts within the punishment environment reported above. No other systematic differences in the effects of personality on trust taking and reciprocity were observed when comparing the strategy method to the direct feedback environment (Table S4). Of note, the model fit (regarding $AICc$) slightly decreased when controlling for the personality x treatment interactions.

S.3.1 Stage 1 and 2: Transfer and back-transfer

Table S4: Regressions investigating the effect of personality on the propensity to trust (T) and back-transfers (BT) in the punishment treatment (PT) compared to the direct feedback (with punishment) treatment (DT). DT is a dummy variable reflective of the presence of direct feedback. The model included a number of control variables which are gender (male = 1), age, as well as culture encoded as being Swiss or not, and city, reflective of living in a city with $>$ 10'000 inhabitants. Constant, treatment-order-dummies, round-one-dummy and session size are not reported in the table.

Model	#1	#2	#1	#2
Dep. Variable	T (Logit)	T (Logit)	BT (OLS)	BT (OLS)
Variable	Coef./(SE)	Coef./(SE)	Coef./(SE)	Coef./(SE)
DT	1.227***	1.256***	2.995***	3.271***
	(0.21)	(0.21)	(0.85)	(0.96)
Male	0.069	0.061	2.878*	2.840*
	(0.43)	(0.44)	(1.51)	(1.50)
City	-1.051**	-1.063**	-0.185	-0.212
	(0.47)	(0.46)	(1.23)	(1.22)
Swiss	-0.107	-0.091	-0.633	-0.777
	(0.44)	(0.44)	(1.25)	(1.24)
Age	0.050	0.051	-0.013	0.007
	(0.07)	(0.07)	(0.12)	(0.12)
MDBF	0.035	0.035	0.289	0.266
	(0.10)	(0.10)	(0.22)	(0.22)
PSS	-0.120**	-0.119**	-0.031	-0.030
	(0.06)	(0.06)	(0.19)	(0.19)
EMO Factor	0.187	0.285	0.462	0.373
	(0.24)	(0.24)	(0.85)	(0.95)
ANTI Factor	0.115	0.086	-1.009	-1.298*
	(0.25)	(0.28)	(0.63)	(0.71)
SS Factor	0.436***	0.460***	0.562	0.426
	(0.17)	(0.17)	(0.59)	(0.73)
ANG Factor	-0.068	-0.066	-1.128*	-1.266*
	(0.15)	(0.16)	(0.63)	(0.70)
IMP Factor	0.356	0.283	0.139	0.010
	(0.27)	(0.30)	(0.47)	(0.57)
EMO x DT	–	-0.298	–	0.298
		(0.19)		(0.70)
ANTI x DT	–	0.087	–	1.143*
		(0.29)		(0.68)
SS x DT	–	-0.041	–	0.457
		(0.16)		(0.78)
ANG x DT	–	-0.018	–	0.630
		(0.18)		(0.78)
IMP x DT	–	0.216	–	0.428
		(0.26)		(0.64)
N (# Clusters)	1080 (90)	1080 (90)	738 (89)	738 (89)
AICc	892	899	5011	5015
(Pseudo) R^2	0.17	0.17	0.13	0.14
Prob. $> \chi^2$	0.00	0.00	0.00	0.00

Significance levels: *** = 1%, ** = 5% and * = 10%.
Robust standard errors, clustered by subjects.

S.4 Relationship between risk attitudes and trust

We investigated the relationship between trust taking and risk attitudes, which has been a recurring concern in the literature (Houser, Schunk & Winter, 2010; Altmann, Dohmen & Wibral, 2008). To dissociate the association between personality and trust taking from potential influences of risk attitudes, we answered the following questions: (1) Is there a general relationship between trust and risk taking?; (2) Is there a relationship between personality and risk attitude as suggested by previous research (Capra et al., 2013) and how does it differ from the relationship between personality and social preferences? If the answer to the former questions was positive, this indicates that trust and risk are somehow interrelated, as well as that personality influences risk attitudes. We therefore asked additionally (3) whether risk attitude can explain the differential behavior under different environments, given that one possible explanation of the observed investment behavior during punishment may be a reduction in the perception of the riskiness of trust due to the ability to punish.

S.4.1 Research methods

After completion of all the trust games, a subset of 104 participants made risky decisions in the context of a certainty equivalent task. The task consisted of a total of 126 individual decisions, in which each choice scenario offered an alternative between choosing a probabilistic lottery and a sure amount. The lottery offered one potential payoff that is greater than the sure amount, and one that is smaller. The payout was determined by randomly selecting 1 of the 126 choice scenarios for which participants earned additional cash amounts between 0 and 50 CHF. Importantly, to reduce the likelihood that knowledge about the risk task influenced trust decisions, subjects were informed of this task only after completion of all trust games. Furthermore, information of trust game winnings was provided after completion of all risk decisions. It is therefore unlikely that decisions on one task influenced decisions on the other.

For the remainder of subjects (N = 78), risk attitudes were assessed via a series of six choices between a lottery (same in all trials: 50 % chance of winning either 10 CHF or 0.5 CHF) and increasing amounts of safe payments (increasing from 2 CHF to 7.5 CHF). The switching point, reflective of the certainty equivalent in this choice scenario, is taken as a measure of risk attitude. The payout was determined by randomly selecting one of the six lotteries for which participants earned additional cash amounts between 0.5 CHF and 10 CHF.

To disentangle trust from risk preferences, we investigated if individual risk parameters (according to the functional form presented in Lattimore, Baker & Witte (1992)) exhibit effects on the propensity to trust, reciprocate and punish, and if personality shows effects on risk parameters, including α (reflective of the curvature of the subjective value function $v(x)$; measures the degree of risk aversion), β (reflective of the elevation of the probability weighting function $w(p)$; measures attractiveness of gambling) and γ (reflective of the curvature

of $w(p)$; measures weighting discriminability), as well as CE (certainty equivalents; reflective of the switching point of a lottery) for a subset of subjects.
(1) We investigated the associations between risk parameters and propensity to trust in all four versions of the trust game, i.e., the NBT, NPT, PT and DT (Table S5). To this end, we conducted regressions that were equivalent to those conducted for stage one of the trust games reported in the main paper, however, with risk parameters as the main independent variables of interest and the propensity to trust as dependent variables.
(2) To address the question whether personality predicts risk attitudes, we examined whether different personality variables show an effect on risk compared to trust taking and reciprocity. We investigated the association between personality and risk parameters using precision-weighted WLS regressions with risk characteristics as dependent variables and personality as independent variables (Table S6).
(3) To investigate whether risk attitudes predict differential trust taking and reciprocity across punishment environments, we run OLS regressions with the propensity to trust and reciprocity as dependent variables and risk parameters as independent variables. Importantly, we included punishment as a predictor variable in all regressions and interacted punishment with risk parameters to investigate the degree to which risk characteristics differentially predict trust game decisions depending on P1s' option to punish or not (Table S7).

S.4.2 Probability weighting and subjective value function

Risk parameters were estimated* for a sub-sample of 104 subjects, where for five individuals, parameters could not be estimated properly (minus 2 outliers from questionnaire task: Total N = 97 subjects). Estimated coefficients represent individual $LBW^{\dagger}$ (Lattimor, Baker & Witte, 1992) weighting function parameters:

$$EV_{Sure,i} = sw^{\alpha_i} \tag{S14}$$

$$EV_{Lottery,i} = (1 - w_i(p_1))z_1^{\alpha_i} + w_i(p_1)z_2^{\alpha_i} \tag{S15}$$

$$\Delta EV_i = EV_{Lottery,i} - EV_{Sure,i} \tag{S16}$$

$$w_i(p_1) = \frac{\beta_i p_1^{\gamma_i}}{\beta_i p_1^{\gamma_i} + (1 - p_1)^{\gamma_i}} \tag{S17}$$

$$P_i(ChooseLottery) = \frac{1}{1 + \exp(-\tau_i \Delta EV_i)} \tag{S18}$$

where in each trial (T = 126 choice scenarios), sw was the sure win, z_1 was the lottery amount smaller than sw and z_2 was the lottery amount larger than

*Maximum likelihood estimation with Newton-Raphson algorithm.

†Note that we compared different weighting functions. The *LBW* function fitted the data best (lowest *AICc*), followed by the *Prelec* weighting function (Prelec, 1998), but the average *AICc* difference was small and insignificant.

sw, p_1 was the winning probability, $w_i(p_1)$ was the weighting function, ΔEV_i was the risk premium and $P_i(ChooseLottery)$ was the probability of choosing the lottery instead of the sure amount. α is reflective of the curvature of the subjective value function $v(x)$, measuring the degree of risk aversion. β is reflective of the elevation of the probability weighting function $w(p)$, measuring the attractiveness of gambling. γ is reflective of the curvature of $w(p)$, measuring the weighting discriminability. τ measures the sensitivity of choice probability to the value difference and, hence, measures the degree of randomness in choice behavior (Hsu et al., 2009). This parameter was not included in later analyses.

S.4.3 Risk task results

The association between LBW risk parameters and propensity to trust in all four versions of the trust game (NBT, NPT, PT, DT) is shown in Table S5. In the two treatments without punishment, β (attractiveness of gambling) exhibited a positive effect on the propensity to trust (NBT: $p < 0.01$; NPT: $p < 0.1$).

We focus our discussion on investor decisions in the non-binary version (NBT) of the trust game, because such a continuous measure of trust taking is a more sensitive measure for investigations concerning the relationship between trust and risk taking. We found a significant relationship between investor trust taking and the intercept parameter (β; $p < 0.01$) of the probability weighting function $w(p)$. This effect is consistent across the trust games without a punishment option (NPT and NBT). This result indicates that greater optimism in the risk domain translates to greater propensity to trust, but only in the absence of punishment, as this effect disappears in those trust games that included a punishment option. This replicates prior observations reported in the literature (e.g. Altmann, Dohmen & Wibral, 2008) that have shown that trust taking and risk attitudes are related.

Table S5: OLS and Logit regressions investigating the effects of individual risk parameters α (subjective value of money; measures the degree of risk aversion), β (elevation; measures the attractiveness of gambling) and γ (curvature; measures the weighting discriminability) on the propensity to trust (T) in all four treatment conditions. The model included a number of control variables which are gender (male = 1), age, as well as culture encoded as being Swiss or not, and city, reflective of living in a city with $>$ 10'000 inhabitants. Constant, treatment-order-dummies, round-one-dummy and session size are not reported in the table.

Treatment Variable	NBT (OLS) Coef./(SE)	NPT (Logit) Coef./(SE)	PT (Logit) Coef./(SE)	DT (Logit) Coef./(SE)
Male	-2.071**	-0.758	0.071	0.441
	(0.82)	(0.49)	(0.58)	(0.74)
City	-0.808	-0.551	-1.807**	-1.291*
	(0.84)	(0.45)	(0.72)	(0.74)
Swiss	1.447	0.432	0.416	-0.259
	(1.24)	(0.62)	(0.67)	(0.85)
Age	-0.002	0.022	0.179*	0.534***
	(0.12)	(0.09)	(0.10)	(0.20)
MDBF	0.240	0.247*	0.204*	0.206
	(0.20)	(0.13)	(0.12)	(0.18)
PSS	-0.149	-0.020	-0.015	-0.162*
	(0.10)	(0.06)	(0.05)	(0.09)
α	2.537	-0.970	-1.119	-2.884
	(2.35)	(1.41)	(1.62)	(2.65)
β	5.038***	2.145*	0.918	-0.140
	(1.56)	(1.13)	(1.48)	(1.58)
γ	2.948	0.085	-0.455	-2.078
	(2.43)	(1.32)	(1.48)	(1.77)
N (# Clusters)		312 (52)		
(Pseudo) R^2	0.37	0.19	0.27	0.28
Prob. $> \chi^2$	0.00	0.16	0.00	0.43

Significance levels: *** = 1%, ** = 5% and * = 10%.
Robust standard errors, clustered by subjects.

The association between personality and LBW risk parameters was investigated using precision-weighted WLS regressions. Only β was significantly affected by personality ($p < 0.01$), showing strong positive effects for sensation seeking and impulsive traits and supporting previous reports of the relationship between personality and risk parameters (Capra et al., 2013). As a higher β-value shifts the weighting curve upwards, such traits tended to increase the attractiveness of gambling. Of note, these results are qualitatively different from the association between personality characteristics and trust and reciprocity as reported in the main paper in Table 2 and Table 4, respectively.

Important for our findings, anti-sociality did not affect risk-taking behavior. This shows that personality has a somewhat dissociable effect in trust and risk taking. Also no other significant associations between risk attitudes and personality were found.[‡] Specifically, personality did not exhibit an effect on

[‡]Higher ANG factor scores tended to lower certainty equivalents (CE), but the overall regression was insignificant ($p > 0.43$).

the curvature of $v(x)$ as well as on the curvature of $w(p)$, as indicated by insignificant regressions investigating the influence of personality on α and γ (p > 0.1; Table S6).

Table S6: Regressions for individual risk and weighting function parameters α (subjective value of money; measures degree of risk aversion), β (elevation; measures the attractiveness of gambling), γ (curvature; measures the weighting discriminability) and CE (certainty equivalent) on personality variables. The model included a number of control variables which are gender (male = 1), age, as well as culture encoded as being Swiss or not, and city, reflective of being from a city $>$ 10'000 inhabitants. Constant and session-dummies are not reported in the table. Note: α, β and γ regressions are based on a $MEMA$ (mixed-effects multilevel analysis) framework according to Chen et al. (2012). Weighted least squares (WLS) regressions accounted for heterogeneous within-subject variability $\hat{\sigma_i}^2$ in estimated risk parameters. $\hat{\tau}^2$ were the $REML$ (restricted maximum likelihood) estimates of between-subject variance.

Subjective value / Weighting function parameter / CE	α	β	γ	CE
Variable	Coef./(SE)	Coef./(SE)	Coef./(SE)	Coef./(SE)
Male	0.065	-0.072	0.064	-0.008
	(0.05)	(0.06)	(0.06)	(0.26)
City	-0.055	0.035	-0.024	0.074
	(0.04)	(0.05)	(0.05)	(0.20)
Swiss	-0.011	-0.093	-0.056	-0.029
	(0.05)	(0.06)	(0.05)	(0.21)
Age	0.008	-0.012	0.011	-0.040*
	(0.01)	(0.01)	(0.01)	(0.02)
MDBF	0.001	-0.022*	0.014	0.004
	(0.01)	(0.02)	(0.01)	(0.04)
PSS	0.004	-0.012*	-0.002	-0.034
	(0.00)	(0.01)	(0.01)	(0.03)
EMO Factor	-0.016	0.021	0.036	0.035
	(0.03)	(0.03)	(0.03)	(0.14)
ANTI Factor	0.008	0.040	-0.013	-0.142
	(0.02)	(0.03)	(0.02)	(0.13)
SS Factor	-0.011	0.107***	0.022	-0.078
	(0.02)	(0.03)	(0.02)	(0.11)
ANG Factor	-0.009	-0.003	0.002	-0.266**
	(0.02)	(0.03)	(0.03)	(0.10)
IMP Factor	-0.022	0.070***	-0.034	0.096
	(0.02)	(0.03)	(0.02)	(0.10)
$\hat{\tau}^2$	0.017	0.036	0.038	–
N	97	97	97	76
R^2	0.17	0.35	0.17	0.14
Prob. $> F$	0.39	0.00	0.55	0.43

Significance levels: *** = 1%, ** = 5% and * = 10%.
Likelihood-ratio test for $\hat{\tau}^2 = 0$: Prob. $> \chi^2 = 0.00$
Weighting matrix: diag. $\left(\frac{1}{\hat{\tau}^2 + \hat{\sigma_i}^2}\right)$

Finally, to address the question whether risk attitudes can explain differential behavior under different punishment environments in the trust game, we

investigated the combined effect of risk parameters and the option to punish on trust taking and reciprocity.

The propensity to trust was not significantly affected by risk parameters (Table S7; note that results focus on reduced models #1 and #3, which both had a lower $AICc$). We found a weak ($p < 0.1$) positive effect of β and a significant positive effect of γ on back-transfers ($p < 0.05$). α showed no significant effect on both the propensity to trust and back-transfer amounts. Importantly, interactions with the treatment dummy PT were insignificant, indicating that risk parameters did not affect trust taking differentially across punishment environments. This latter point was further affirmed by the findings that the reduced models without interaction terms provided better model fits.[§]

S.4.4 Risk task conclusions

Results indicate that individual risk attitudes did not have an impact on the relationships between trust and personality, as well as reciprocity and personality because (1) the personality variables that showed a relationship with risk parameters were different from those that exhibited an association with social choice parameters and (2) risk attitudes do not explain the changes in behavior in the presence relative to the absence of punishment. In all, our results indicate that decisions to trust and to reciprocate in anonymous one-shot interactions were only marginally explained by risk taking characteristics (Fehr, 2009; Houser, Schunk & Winter, 2010). Importantly, risk attitudes did not account for behavioral changes in the punishment environment.

[§]Note: We also investigated the whether risk attitudes measured by certainty equivalents (CE) predict trust taking across punishment environments. We did not find an effect of CEs on the decision to transfer 10 CHF, as well as on back-transfers. Importantly, interactions of CEs with the treatment dummy PT did not affect trust game outcomes either.

Table S7: Regressions investigating the effect of risk and weighting function parameters α (subjective value of money; measures degree of risk aversion), β (elevation; measures the attractiveness of gambling) and γ (curvature; measures the weighting discriminability) on the propensity to trust (T) and back-transfers (BT) in the presence compared to the absence of the option to punish. PT is a dummy variable reflective of the option to punish. The model included a number of control variables which are gender (male = 1), age, as well as culture encoded as being Swiss or not, and city, reflective of living in a city with $>$ 10'000 inhabitants. Constant, treatment-order-dummies, round-one-dummy and session size are not reported in the table.

Model	#1	#2	#3	#4
Dep. Variable	T (Logit)	T (Logit)	BT (OLS)	BT (OLS)
Variable	Coef./(SE)	Coef./(SE)	Coef./(SE)	Coef./(SE)
PT	0.757***	1.378	5.856***	11.537*
	(0.22)	(1.09)	(1.30)	(6.22)
Male	-0.708	-0.710	-3.210**	-3.210**
	(0.47)	(0.47)	(1.52)	(1.53)
City	-0.578	-0.581	-0.089	-0.089
	(0.55)	(0.55)	(1.94)	(1.95)
Swiss	0.356	0.347	0.659	0.659
	(0.66)	(0.66)	(1.74)	(1.75)
Age	0.109	0.109	0.752**	0.752**
	(0.08)	(0.08)	(0.31)	(0.31)
MDBF	0.263**	0.263**	-0.201	-0.201
	(0.12)	(0.12)	(0.32)	(0.32)
PSS	0.004	0.004	0.067	0.067
	(0.05)	(0.05)	(0.14)	(0.14)
α	-1.095	-0.901	4.981	7.977
	(1.62)	(1.69)	(3.94)	(5.19)
β	1.852	2.216*	4.809*	6.920*
	(1.21)	(1.23)	(2.67)	(3.90)
γ	-1.229	-1.355	10.072**	7.790
	(1.50)	(1.50)	(3.83)	(4.82)
α x PT	–	-0.394	–	-5.992
		(0.98)		(5.71)
β x PT	–	-0.795	–	-4.222
		(0.86)		(3.99)
γ x PT	–	0.300	–	4.565
		(1.46)		(4.93)
N (# Clusters)	624 (52)	624 (52)	540 (45)	540 (45)
AICc	639	654	3698	3710
(Pseudo) R^2	0.21	0.21	0.36	0.37
Prob. $> \chi^2$	0.00	0.00	0.00	0.00

Significance levels: *** = 1%, ** = 5% and * = 10%.
Robust standard errors, clustered by subjects.

References

Altmann, S., Dohmen, T. & Wibral, M. (2007). Do the Reciprocal Trust Less? *Economic Letters*, 99, 454-457.

Burnham, K. P. & Anderson, D. (2002). *Model Selection and Multi-Model Inference*. Springer (2nd edition).

Capra, M., Jiang, B., Engelmann, J. B. & Berns, G. (2013). Can Personality Type Explain Heterogenety in Probability Distortions? *Journal of Neuroscience, Psychology, and Economics*, 6(3), 151-166.

Chen, G., Saad, Z. S., Beauchamp, M. S. & Cox, R. W. (2012). FMRI Group Analysis Combining Effect Estimates and their Variances. *NeuroImage*, 60, 747-765.

Fehr, E. (2009). On the Economics and Biology of Trust. *Journal of the European Economic Association*, 7(2-3), 235-266.

Houser, D., Schunk, D. & Winter, J. (2010). Distinguishing Trust from Risk: An Anatomy of the Investment Game. *Journal of Economic Behavior & Organization*, 74, 72-81.

Hsu, M., Krajbich, I., Zhao, C. & Camerer, C. F. (2009). Neural Response to Reward Anticipation Under Risk is Nonlinear in Probabilities. *The Journal of Neuroscience*, 29(7), 2231-2237.

Kable, J. W. & Glimcher, P. W. (2007). The Neural Correlates of Subjective Value during Intertemporal Choice. *Nature Neuroscience*, 10(12), 1625-1633.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2), 443-478.

Lattimore, P. K., Baker, J. R. & Witte, A. D. (1992). The Influence of Probability on Risky Choice: A Parametric Examination. *Journal of Economic Behavior & Organization*, 17(3), 377-400.

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, 66(3), 497-527.

Wagenmakers, E. J. & Farrell, S. (2004). AIC Model Selection using Akaike Weights. *Psychonomic Bulletin & Review*, 11(1), 192-196.